2016

Modern Estimation Problems in Group Testing

Md Shamim Sarker
University of South Carolina

Modern estimation problems in group testing

by

Md Shamim Sarker

Bachelor of Science
Jahangirnagar University 2005

Master of Science
Lamar University 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Statistics

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

Joshua M. Tebbs, Major Professor

Timothy Hanson, Committee Member

Lianming Wang, Committee Member

Alexander C. McLain, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Dedication

I dedicate this dissertation to my loving family. Their dedication, love, and patience have guided and encouraged me to advance every single step of this long journey. I would also like to dedicate this work to the University of South Carolina, where I have spent the most valuable time and have received the most rewarding experience of my life.

Acknowledgments

First of all, I would like to express my heartfelt gratitude to my advisor Dr. Joshua Tebbs for his expert guidance, unwavering support, and encouragement throughout my study at the University of South Carolina. He has continually conveyed a spirit of adventure towards research and scholarship. Without his incredible patience and timely direction, my dissertation would have been a frustrating and overwhelming pursuit. His teachings and the valuable experiences he has shared with me will always accompany me along the academic career I am going to begin. Dr. Tebbs is a mentor, colleague, and friend who I will eternally cherish.

I would like to express my heartfelt appreciation to other committee members. I acknowledge the advice Dr. Hanson has offered throughout the process of this dissertation. His insightful comments have shaped my thoughts to play with new research ideas and to solve many complicated problems. I would like to thank Dr. Lianming Wang for his valuable contribution to this work. I also wish to acknowledge the value of Dr. Wang's courses, as they have had a direct positive impact on my research. I would like to warmly thank Dr. Alexander McLain for his insightful comments, which directed me to think about my work more thoroughly.

I would like to thank Dr. Xianzheng Huang who has taught me a number of courses and with whom I have worked in collaboration. I have had the opportunity to share my ideas with her and receive valuable directions. The measurement error project presented in Chapter 4 is motivated by a course that Dr. Huang taught. I would like to thank other faculty members who were more than generous with their expertise and precious time. I am thankful to Ms. Maureen Petkewich for her

continued support and advice on teaching and many administrative activities. I am very thankful to the staff in the Department of Statistics for their support in various administrative affairs. I would also like to acknowledge the help from faculty members and staff from other departments.

I have been privileged to have Dr. Christopher McMahan as a collaborator. His unwavering support, critical assessment of my work, and very insightful suggestions have had a tremendous impact on development of my dissertation. I would also like to thank Dr. Christopher Bilder for his diligent effort and thoughtful advice. I am extremely thankful to my M.S. advisor Dr. Kumer Pial Das who has always been a mentor and a true friend. Finally, I express my heartfelt appreciation to my friends, fellow graduate students, and many people who have walked alongside me.

Abstract

In the simplest form of group testing, pools are formed by compositing a fixed number of individual specimens (e.g., blood, urine, swab, etc.) and then the pools are tested for a binary characteristic, such as presence or absence of a disease. Group testing is commonly used to screen for a variety of sexually transmitted diseases in epidemiological applications where the main goal is to increase testing efficiency. In this dissertation, we study three estimation problems that are motivated by real-life applications. We propose new methods to model group testing data for both single and multiple infections. In the first problem, we propose a Bayesian approach to estimate the prevalence of multiple infections. This relaxes the unreliable assumption that diagnostic accuracies are constant. Also, when historical data are taken into account, our method provides more efficient estimation than do existing approaches. In the second problem, we propose a regression method to capture dilution effects due to pooling. In addition to offering reliable inference, our parametric approach enables one to perform a hypothesis test for dilution. In the third problem, we propose Bayesian measurement error models. Our approach provides flexibility to the structural modeling approach which requires the availability of a known probability distribution for true (unobserved) covariates. This work generalizes existing regression methods to account for covariate measurement error. We also discuss several problems for future research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Literature review

We start with a brief literature review of group testing classification (case identification) and estimation problems. Before starting our main discussion, we define two accuracy measures of a diagnostic assay that are associated with group testing. The sensitivity is the probability that a positive sample is diagnosed as positive, and the specificity is the probability that a negative sample is diagnosed as negative. In a multiple-infection problem, we define these accuracy measures for each infection separately (see Chapter 2). To capture pooled dilution effects, we model the sensitivity in Chapter 3 as an increasing function of the number of true positives within a pool.

Background

With the goal of minimizing testing costs, Dorfman (1943) introduced the idea of group testing to screen US soldiers for syphilis during the Second World War. Dorfman's method, which is commonly referred to as "Dorfman testing," is a two-stage hierarchical algorithm where initial (non-overlapping) master pools are tested in stage 1 and then individual retesting is performed in stage 2. If a pool is diagnosed as negative, all individuals in the pool are declared negative; on the other hand, if a pool is diagnosed as positive, all individuals in the pool are retested one-by-one. As demonstrated by Dorfman, this simple two-stage algorithm can offer substantial savings in testing costs when the infection rate is small. Since Dorfman's seminal work, group

testing has been widely accepted as a cost-effective alternative to individual testing in many applications, including public health, genetics, animal disease testing, drug discovery, and pollution detection.

Classification

The use of Dorfman's two-stage procedure has been well regarded by both practitioners and researchers. The popularity of this method can likely be explained by the fact that it is simple to implement. Many variations of Dorfman's algorithm are currently available. For example, Pilcher et al. (2005) uses a three-stage algorithm which involves a second stage of testing subpools if the master pool tests positively. Halving algorithm (Litvak et al., 1994) involves multiple stages in which each pool that tests positively is split further into two halves before testing is performed in subsequent stages. The final stage of halving involves individual testing.

Unlike hierarchical algorithms, non-hierarchical algorithms use overlapping pools. An example of a non-hierarchical algorithm is square array testing (Phatarfod and Sudbury, 1994; Kim et al., 2007). Before performing initial tests, n^2 specimens are placed in an $n \times n$ matrix. Then n pools are formed by taking samples from n rows and, similarly, another n pools are created from samples of n columns. These $2n$ pools are then tested. Specimens at the intersection of a positive row and a positive column are tested individually. A more generalized square array testing algorithm allows for the possibility of testing errors. Kim et al. (2007) summarizes the operating characteristics (i.e., classification efficiency and accuracy) of hierarchical and square array testing algorithms in the presence of testing error.

Sterrett (1957) proposed an extension of Dorfman's procedure whereby individuals from positive pools are retested in multiple stages. According to this procedure, individuals in a positive pool are tested one-by-one with random selection until the first positive individual is identified. The remaining individuals are then used to form

a new pool. If the new pool tests negatively, all individuals are declared negative; however, if the new pool tests positively, individuals in the new pool are again tested randomly until the first positive one is identified. This procedure is repeated until all subjects are classified as positive or negative. Sterrett's decoding procedure can be more efficient than Dorfman's when the probability of infection is small.

Sterrett's procedure has been improved upon by Bilder et al. (2010) who used individuals' covariate information to perform informative retesting. In this approach, each individual's likelihood of disease positivity is first estimated using covariate information. When individuals from positive pools are to be retested, an individual who is most likely to be positive is tested first. Motivated by this informative approach, McMahan et al. (2012a) generalized array testing to account for the heterogeneity among individuals. McMahan et al. (2012b) suggested an informative version of Dorfman decoding in which pools are formed based on individuals' risk probabilities. Informative approaches can be significantly more efficient when compared to the corresponding non-informative approaches; i.e., those that do not account for heterogeneity among individuals.

As testing for multiple infections is becoming more common, group testing research is also shifting. This is because recently developed assays can accurately detect multiple infections simultaneously from a single specimen. Tebbs et al. (2013) first studied a two-stage Dorfman-type testing protocol adopted by the Infertility Prevention Project (IPP) for chlamydia and gonorrhea testing. These authors demonstrated that efficiency can be increased dramatically when pool testing involves multiple infections. The findings presented in Tebbs et al. (2013) serve as motivation to extend existing classification algorithms from single to multiple infections. Further advancements can be possible when exploiting heterogeneity among individuals as proposed by Bilder et al. (2010) for single infections.

Estimation

While group testing research for classification has flourished over the past decades, estimation has also received substantial attention. Testing data obtained from any group testing protocol, such as Dorfman decoding, halving, and array testing, can be modeled to estimate either an overall disease prevalence or individual-level disease probabilities using covariates. For a rare disease, inference based on group testing can be as efficient as that for individual testing at only a fraction of the testing cost. Even though retesting individuals from positive pools is necessary for classification purposes, retesting is not crucial for estimation. This can be explained heuristically as follows. If a disease is rare, pools that test positively may not contain more than one positive case; hence, retesting individuals from positive pools provides little additional information. Therefore, the majority of group testing papers that proposed estimation methods used testing responses from only initial master pools. The advantages of this approach include simplicity in statistical modeling and additional reductions in testing costs.

Most researchers in group testing, until the late 1990's, focused only on estimating the proportion (overall prevalence) of a rare binary trait. The first such estimation paper is Thompson (1962), who took a maximum likelihood (ML) approach to estimate the proportion of insect vectors capable of transmitting aster-yellow virus in a population of aphids. Thompson's work proceeded under the assumption that testing results are perfect and that insects' infection statuses are independent. In addition to using master pool responses as in Thompson (1962), Sobel and Elashoff (1975) allowed the statistical model to incorporate retest information. Hwang (1976) proposed a maximum likelihood estimator (MLE) in the presence of dilution effects. Burrows (1987) took an alternative ML approach which can improve estimation in terms of bias and efficiency. Robustness of estimation has been studied by several authors including Chen and Swallow (1990) and Hung and Swallow (1999). A number of

authors have studied group testing optimality; among others, see Tu et al. (1995) and Liu et al. (2012).

Group testing estimation from within a Bayesian paradigm has also been studied. Chaubey and Li (1995) presented a Bayesian method for estimation and showed that Bayesian estimators can be preferred to ML estimators. Mendoza-Blanco et al. (1996) presented a general Bayesian framework that models data resulting from a variety of sampling strategies. Bilder and Tebbs (2005) proposed an empirical Bayes method for estimation using master pools only. Johnson and Pearson (1999) developed a Bayesian methodology for a two-stage testing where individuals from pools that test negatively are re-pooled. This technique was originally proposed by Gastwirth and Johnson (1994) who took a frequentist approach. A similar method was proposed by Hanson et al. (2006), who acknowledge heterogeneity among populations due to regional differences and allow the model to incorporate varying prevalences.

With the exception of Hanson et al. (2006), all of the estimation methods discussed above aim at estimating a single proportion without accounting for heterogeneity among individuals. Recent work has focused on developing regression methodology using covariates, such as age, gender, and disease symptoms, to obtain individual-level estimates. Farrington (1992) first proposed a regression method for a specific generalized linear model with the stringent assumption that each individual within a pool shares identical covariates. Vansteelandt et al. (2000) extended this work to allow for any type of covariate structure. Xie (2001) presented a general expectation-maximization methodology which can incorporate retest results. Bilder and Tebbs (2009) studied estimation bias and efficiency for regression estimates using the model introduced by Vansteelandt et al. (2000). Huang and Tebbs (2009) and Huang (2009) developed diagnostic methods to identify latent model misspecification for structural measurement error models using group testing responses. McMahan et al. (2013) proposed a regression approach that accounts for pooled dilution effects. Wang et al.

(2015) generalized the regression method of McMahan et al. (2013) to be applicable for any group testing algorithm.

A number of papers have presented nonparametric regression estimation methods for group testing data. The seminal work in Delaigle and Meister (2011) used test results from randomly formed pools. Delaigle and Hall (2012) later presented a nonparametric approach for pools formed homogeneously. Wang et al. (2014b) presented a general semiparametric approach to model data from any group testing algorithm. Delaigle and Zhou (2015) presented a nonparametric extension of the dilution methods in McMahan et al. (2013).

Switching gears to multiple infections, Hughes-Oliver and Rosenberger (2000) first proposed a method to estimate the prevalence of multiple diseases with the assumption that a perfect assay test is available to detect all diseases simultaneously. Tebbs et al. (2013) extended this work to allow for imperfect testing and to incorporate individual retest results. Zhang et al. (2013) proposed a regression method for multiple diseases. The research direction of group testing with multiple diseases is becoming popular because of its additional cost savings and also because of the availability of multiple-infection assays.

Subsequent chapters of this dissertation are organized as follows. In Chapter 2, we present a Bayesian model to estimate the prevalence of multiple infections. In Chapter 3, we present a regression method for single traits that accounts for dilution. In Chapter 4, we propose a Bayesian framework that corrects for measurement errors in covariates. Finally, we describe future research ideas in Chapter 5. Supplementary materials from Chapters 2 and 3 are provided in Appendices A and B.

Chapter 2

Estimating the prevalence of multiple diseases from two-stage hierarchical pooling[1]

The material in this chapter and Appendix A are taken from the manuscript, "Estimating the prevalence of multiple diseases from two-stage hierarchical pooling," by M. Warasi, J. Tebbs, C. McMahan, and C. Bilder. This manuscript was accepted at *Statistics in Medicine* on 03/17/2016. Permission to reprint is shown in Appendix C.

Summary: Testing protocols in large-scale sexually transmitted disease screening applications often involve pooling biospecimens (e.g., blood, urine, swabs, etc.) to lower costs and to increase the number of individuals who can be tested. With the recent development of assays that detect multiple diseases, it is now common to test biospecimen pools for multiple infections simultaneously. Recent work has developed an expectation-maximization algorithm to estimate the prevalence of two infections using a two-stage, Dorfman-type testing algorithm motivated by current screening practices for chlamydia and gonorrhea in the United States. In this article, we have the same goal but instead take a more flexible Bayesian approach. Doing so allows us to incorporate information about assay uncertainty during the testing process, which involves testing both pools and individuals, and also to update information as individuals are tested. Overall, our approach provides reliable inference for disease probabilities and accurately estimates assay sensitivity and specificity even when little

[1]Warasi, M., Tebbs, J., McMahan, C., and Bilder, C. (2016). Estimating the prevalence of multiple diseases from two-stage hierarchical pooling. *Statistics in Medicine*, in press. DOI: 10.1002/sim.6964. Reprinted here with permission of John Wiley and Sons.

or no information is provided in the prior distributions. We illustrate the performance of our estimation methods using simulation and by applying them to chlamydia and gonorrhea data collected in Nebraska.

2.1 Introduction

Testing biospecimens in pools, which is known as group testing (or pooled testing), is a cost-effective alternative to individual testing in a variety of disease screening applications. Originally proposed by Dorfman (1943) to screen World War II soldiers for syphilis, group testing is now widely used to screen human populations for sexually transmitted diseases, including HIV (Pilcher et al., 2005), HBV and HCV (Hourfar et al., 2008; Stramer et al., 2013), and chlamydia and gonorrhea (Lindan et al., 2005), and for other infectious diseases including West Nile virus (Busch et al., 2005), malaria (Wang et al., 2014a), and influenza (Van et al., 2012). Group testing also arises in other applications, including drug discovery (Remlinger et al., 2006), genetics (Chi et al., 2009), animal disease testing (Dhand et al., 2010), and food safety (Fahey et al., 2006).

Because pooling has become so widespread, statistical research in group testing has also flourished. This research has generally followed two different paths. In the classification (case identification) problem, the goal is to classify each individual as positive or negative. This involves retesting individuals in pools that test positively; see Kim et al. (2007) for a review. In the estimation problem, responses from pools provide enough information to estimate a population prevalence, at times, more efficiently than when individual testing is used (Liu et al., 2012; Zhang et al., 2013). Recent work has focused on the development of regression methods to estimate subject-specific probabilities, either parametrically (Vansteelandt et al., 2000; Chen et al., 2009), semi-parametrically (Wang et al., 2014b), or non-parametrically (Delaigle and Meister, 2011).

This article is motivated by screening practices for chlamydia and gonorrhea (CT/NG) in the United States as part of a national program formerly known as the Infertility Prevention Project (IPP). Chlamydia and gonorrhea are two of the most common sexually transmitted diseases; together, there are approximately 1.5 million new infections reported each year in the United States (Gaydos et al., 2010). The IPP was a federally funded program managed by the Centers for Disease Control and Prevention (CDC) and implemented in each of the 50 states during 1988-2013. After the Affordable Care Act (ACA) was passed in 2010 and implemented in 2014, screening for CT/NG has continued in each state but now testing centers rely on other sources of funding (e.g., federal health care plans, private insurance, etc.). To reduce costs while still screening the same number of individuals for CT/NG, Iowa's IPP program switched from individual testing to group testing in 1999. Doing so has led to millions of dollars in savings (Jirsa, 2008), and other states (Lewis et al., 2012) have since adopted group testing as well. In light of new funding uncertainties created by the ACA (JSI Research & Training Institute, 2015), Iowa's application of group testing might serve as a model for how to perform CT/NG screening nationwide.

Estimating the prevalence of a single disease has received a large amount of attention in the group testing literature. However, testing procedures for CT/NG and other infections are now moving towards the use of assays which detect multiple infections at once (Gaydos et al., 2010). In these instances, pools of individuals are tested for multiple infections using a single assay, and then pools are resolved (decoded) for each infection. Estimation in this situation is challenging, because the true infection statuses on the same individual are latent (due to inherent assay error) and are also correlated. Recently, Tebbs et al. (2013) developed an expectation-maximization (EM) algorithm to jointly estimate the prevalence of CT/NG, motivated by screening practices in Iowa which use group testing (see Section 2.2). Their work, in the two-infection case, generalized the estimation approach in Hughes-Oliver and Rosenberger

(2000) to allow for assay error and also for the inclusion of retesting information on positive pools.

In this article, we have the same estimation goals as in Tebbs et al. (2013), but we take a Bayesian approach instead. Doing so confers important advantages. First, it allows us to relax the potentially untrustworthy assumption that diagnostic test accuracy rates (i.e., sensitivity and specificity) are fixed and known. In practice, these rates are usually estimated on the basis of small pilot studies that manufacturers publish in their product literature. Ignoring the variability in these estimates could compromise inference, especially if the estimates deviate substantially from the true accuracy rates and/or if assay performance varies according to other factors (CDC, 2015). Second, a Bayesian approach is natural given the sequential manner in which screening data amass over time. For example, the State Hygienic Laboratory (SHL) in Iowa City has screened thousands of Iowa residents each year for CT/NG, dating back to 1992. This affords investigators ample information to construct sensible prior distributions as well as to periodically update information on disease prevalence and assay performance. Our work extends previous Bayesian group testing estimation approaches for single diseases (Johnson and Pearson, 1999; Hanson et al., 2006).

Subsequent sections of this article are organized as follows. In Section 2.2, we describe the screening algorithm for CT/NG used in Iowa. This two-stage algorithm was described in detail in Tebbs et al. (2013), so we herein summarize only the salient aspects. In Section 2.3, we present our estimation methods and discuss prior model selection. In Section 2.4, we use simulation to assess estimation performance under a variety of prior models, including models which incorporate little or no information about disease prevalence and assay accuracy. In Section 2.5, we analyze IPP data in the same manner as in Tebbs et al. (2013) to illustrate the advantages of estimation from a Bayesian point of view. In Section 2.6, we conclude with a brief summary discussion. We use Appendix A to show how one could generalize our work to estimate

probabilities for more than two infections if needed.

2.2 Two-stage pooling algorithm

The estimation methods we develop in this article are motivated by the two-stage pooling algorithm described below. This algorithm is used to complete CT/NG testing at the SHL in Iowa City and is potentially applicable in other situations.

POOLING ALGORITHM

Stage 1: Individuals are randomly assigned to master pools. Each pool is tested for both infections using a single assay. A single assay detects both infections simultaneously.

Stage 2: Individuals in pools that

- test negatively for both infections are diagnosed as negative for both infections.
- test positively for either infection are retested (individually) for both infections using the same assay in Stage 1. Diagnoses for both infections are made from the outcomes of the individual tests.

Tebbs et al. (2013) describe various logistical issues of this pooling procedure (as it relates to implementation at the SHL) that we do not repeat here. The point worth emphasizing is that, for simplicity, the SHL uses one assay, the Aptima Combo 2 Assay (Hologic/Gen-Probe, Inc., San Diego) nucleic acid amplification test, for its CT/NG testing. This assay detects both infections simultaneously when it is applied to pools (in Stage 1) and to individuals (in Stage 2). In the infectious disease testing literature, such an assay is said to *discriminate* because it elicits a diagnosis for each infection separately. In this article, we assume that a discriminating assay is available and that it can be applied to both pooled and individual specimens. The literature is replete with examples of multiple-infection assays that are discriminating in nature.

For example, most assays based on nucleic acid amplification technology used for CT/NG detection discriminate between the two infections in both urine and swab specimens (Gaydos et al., 2010; CDC, 2015). Furthermore, the CDC recommends that nucleic acid amplification testing be used for laboratory-based CT/NG detection (CDC, 2015).

2.3 Bayesian estimation

Model formulation and inference

Suppose N individuals are to be tested for two infections (e.g., CT/NG, etc.) using the algorithm described in Section 2.2. Let $\widetilde{\mathbf{Y}}_{ik} = (\widetilde{Y}_{i1k}, \widetilde{Y}_{i2k})'$ denote the vector of true individual binary statuses, for $i = 1, 2, ..., c_k$ and $k = 1, 2, ..., K$, where $N = \sum_{k=1}^{K} c_k$. We call c_k the pool size for the kth master pool. The number of master pools formed at Stage 1 is K. We assume the $\widetilde{\mathbf{Y}}_{ik}$'s, conditional on $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, are independent and identically distributed random vectors with probability mass function

$$\text{pr}(\widetilde{Y}_{i1k} = \widetilde{y}_1, \widetilde{Y}_{i2k} = \widetilde{y}_2|\mathbf{p}) = p_{00}^{(1-\widetilde{y}_1)(1-\widetilde{y}_2)} p_{10}^{\widetilde{y}_1(1-\widetilde{y}_2)} p_{01}^{(1-\widetilde{y}_1)\widetilde{y}_2} p_{11}^{\widetilde{y}_1\widetilde{y}_2},$$

where $\widetilde{y}_1, \widetilde{y}_2 \in \{0, 1\}$ and $p_{00} + p_{10} + p_{01} + p_{11} = 1$. Note that because of inherent assay error, the $\widetilde{\mathbf{Y}}_{ik}$'s are best regarded as latent.

Let $\widetilde{\mathbf{Z}}_k = (\widetilde{Z}_{1k}, \widetilde{Z}_{2k})'$ denote the vector of true binary statuses for the kth master pool, where $\widetilde{Z}_{jk} = I(\sum_{i=1}^{c_k} \widetilde{Y}_{ijk} > 0)$, for $j = 1, 2$, and $I(\cdot)$ is the indicator function. In other words, $\widetilde{Z}_{jk} = 1$ if at least one individual in the kth master pool is truly positive for the jth infection, $\widetilde{Z}_{jk} = 0$ otherwise. Let $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ denote the vector of testing responses observed for the kth master pool in Stage 1, where $Z_{jk} = 1$ if the kth master pool tests positively for the jth infection, $Z_{jk} = 0$ otherwise. If the kth master pool tests positively for at least one infection in Stage 1, let $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})'$ denote the vector of individual testing responses observed for the ith individual,

$i = 1, 2, ..., c_k$, in Stage 2. We allow for pools (in Stage 1) and individuals (in Stage 2) to be misclassified and denote the assay sensitivity and specificity by

$$\begin{aligned} S_{e:j} &= \text{pr}(Z_{jk} = 1|\widetilde{Z}_{jk} = 1) = \text{pr}(Y_{ijk} = 1|\widetilde{Y}_{ijk} = 1) \\ S_{p:j} &= \text{pr}(Z_{jk} = 0|\widetilde{Z}_{jk} = 0) = \text{pr}(Y_{ijk} = 0|\widetilde{Y}_{ijk} = 0), \end{aligned}$$

respectively, for $j = 1, 2$. We assume $S_{e:j}$ and $S_{p:j}$ do not depend on the pool size c_k in Stage 1 so that these probabilities also apply for individual tests performed in Stage 2. This assumption is common in group testing research for single infections (Kim et al., 2007). For this to be reasonable in practice, assay detection thresholds and/or dilution ratios may need to be changed to accommodate both pooled and individual specimens; see McMahan et al. (2013) and the references therein.

The observed data from the pooling algorithm in Section 2.2 consist of (a) the testing responses $\mathbf{Z}_k = (Z_{1k}, Z_{2k})'$ from the K master pools in Stage 1 and (b) the additional c_k individual testing responses $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k})'$ from those pools which tested positively for either infection in Stage 1. For notational purposes, we aggregate all master pool testing responses into a vector denoted by $\mathbf{Z}$ and all individual testing responses into a vector denoted by $\mathbf{Y}$. Let $\boldsymbol{\theta} = (\mathbf{p}', \boldsymbol{\delta}')'$, where $\boldsymbol{\delta} = (S_{e:1}, S_{e:2}, S_{p:1}, S_{p:2})'$. Although it is possible to write out the observed data likelihood $\pi(\mathbf{Z}, \mathbf{Y}|\boldsymbol{\theta})$, its form is not easily amenable to performing a Bayesian analysis. Therefore, we use a data augmentation step that introduces the individuals' true infection statuses $\widetilde{Y}_{ijk}$ as latent random variables. Let $\widetilde{\mathbf{Y}}$ denote the vector that aggregates all of the latent $\widetilde{Y}_{ijk}$'s. The joint distribution of the observed data $\{\mathbf{Z}, \mathbf{Y}\}$ and the latent data $\widetilde{\mathbf{Y}}$, conditional

on $\boldsymbol{\theta}$, can be expressed as

$$\pi(\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}|\boldsymbol{\theta}) = \prod_{k=1}^{K}\prod_{i=1}^{c_k} p_{00}^{(1-\widetilde{Y}_{i1k})(1-\widetilde{Y}_{i2k})} p_{10}^{\widetilde{Y}_{i1k}(1-\widetilde{Y}_{i2k})} p_{01}^{(1-\widetilde{Y}_{i1k})\widetilde{Y}_{i2k}} p_{11}^{\widetilde{Y}_{i1k}\widetilde{Y}_{i2k}}$$
$$\times \left[\prod_{j=1}^{2}\prod_{k=1}^{K} \left(S_{e:j}^{Z_{jk}} \overline{S}_{e:j}^{1-Z_{jk}}\right)^{I(\sum_{i=1}^{c_k}\widetilde{Y}_{ijk}>0)} \left(S_{p:j}^{1-Z_{jk}} \overline{S}_{p:j}^{Z_{jk}}\right)^{I(\sum_{i=1}^{c_k}\widetilde{Y}_{ijk}=0)} \right.$$
$$\left. \times \left\{ \prod_{i=1}^{c_k} S_{e:j}^{Y_{ijk}\widetilde{Y}_{ijk}} \overline{S}_{e:j}^{(1-Y_{ijk})\widetilde{Y}_{ijk}} S_{p:j}^{(1-Y_{ijk})(1-\widetilde{Y}_{ijk})} \overline{S}_{p:j}^{Y_{ijk}(1-\widetilde{Y}_{ijk})} \right\}^{I(Z_{+k}>0)} \right], \quad (2.1)$$

where $Z_{+k} = Z_{1k} + Z_{2k}$, $\overline{S}_{e:j} = 1 - S_{e:j}$, and $\overline{S}_{p:j} = 1 - S_{p:j}$. The first line in Equation (2.1) represents the contribution of the individual latent statuses, while the part within the brackets describes the contributions from Stage 1 (master pool test results; second line) and Stage 2 (individual test results; third line). Note that if $\boldsymbol{\delta}$ were known, Equation (2.1) would be the same as the complete data likelihood in Tebbs et al. (2013). For further discussion on additional assumptions underpinning the construction of $\pi(\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}|\boldsymbol{\theta})$ in Equation (2.1), see Section 2.6.

To complete our Bayesian model specification, we elicit independent beta prior distributions for the assay test accuracies; i.e., $S_{e:j} \sim \text{beta}(a_{S_{e:j}}, b_{S_{e:j}})$ and $S_{p:j} \sim \text{beta}(a_{S_{p:j}}, b_{S_{p:j}})$, for $j = 1, 2$, where all hyperparameters are known. For the vector of infection status probabilities, we specify a Dirichlet prior; i.e.,

$$\mathbf{p} \sim \pi(\mathbf{p}) = B(\boldsymbol{\alpha}) p_{00}^{\alpha_{00}-1} p_{10}^{\alpha_{10}-1} p_{01}^{\alpha_{01}-1} p_{11}^{\alpha_{11}-1},$$

where $B(\boldsymbol{\alpha})$ is a normalizing constant and $\boldsymbol{\alpha} = (\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11})'$ is a vector of known hyperparameters. We assume the test accuracies $S_{e:j}$ and $S_{p:j}$ are both independent of $\mathbf{p}$, for $j = 1, 2$. These assumptions are analogous to the assumptions made in Johnson and Pearson (1999) and Hanson et al. (2006) for single infections.

With these prior choices and assumptions, the "full conditional" distributions can be easily derived from the augmented likelihood function $\pi(\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}|\boldsymbol{\theta})$. For the assay accuracies, these distributions are given by

$$\begin{aligned} S_{e:j}|\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}} &\sim \text{beta}(a^*_{S_{e:j}}, b^*_{S_{e:j}}) \\ S_{p:j}|\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}} &\sim \text{beta}(a^*_{S_{p:j}}, b^*_{S_{p:j}}), \end{aligned}$$

for $j = 1, 2$, where

$$\begin{aligned} a^*_{S_{e:j}} &= a_{S_{e:j}} + \sum_{k=1}^{K} \left\{ Z_{jk}\widetilde{Z}_{jk} + I(Z_{+k} > 0) \sum_{i=1}^{c_k} Y_{ijk}\widetilde{Y}_{ijk} \right\} \\ b^*_{S_{e:j}} &= b_{S_{e:j}} + \sum_{k=1}^{K} \left\{ (1 - Z_{jk})\widetilde{Z}_{jk} + I(Z_{+k} > 0) \sum_{i=1}^{c_k} (1 - Y_{ijk})\widetilde{Y}_{ijk} \right\} \\ a^*_{S_{p:j}} &= a_{S_{p:j}} + \sum_{k=1}^{K} \left\{ (1 - Z_{jk})(1 - \widetilde{Z}_{jk}) + I(Z_{+k} > 0) \sum_{i=1}^{c_k} (1 - Y_{ijk})(1 - \widetilde{Y}_{ijk}) \right\} \\ b^*_{S_{p:j}} &= b_{S_{p:j}} + \sum_{k=1}^{K} \left\{ Z_{jk}(1 - \widetilde{Z}_{jk}) + I(Z_{+k} > 0) \sum_{i=1}^{c_k} Y_{ijk}(1 - \widetilde{Y}_{ijk}) \right\} \end{aligned}$$

and $\widetilde{Z}_{jk} = I(\sum_{i=1}^{c_k} \widetilde{Y}_{ijk} > 0)$. For the prevalence parameter $\mathbf{p}$, the full conditional distribution is again Dirichlet; i.e., $\mathbf{p}|\widetilde{\mathbf{Y}} \sim \text{Dirichlet}(\boldsymbol{\Psi})$, where

$$\begin{aligned} \boldsymbol{\Psi} = \Bigg(&\alpha_{00} + \sum_{k=1}^{K}\sum_{i=1}^{c_k} \widetilde{V}_{(00)ik}, \alpha_{10} + \sum_{k=1}^{K}\sum_{i=1}^{c_k} \widetilde{V}_{(10)ik}, \\ &\alpha_{01} + \sum_{k=1}^{K}\sum_{i=1}^{c_k} \widetilde{V}_{(01)ik}, \alpha_{11} + \sum_{k=1}^{K}\sum_{i=1}^{c_k} \widetilde{V}_{(11)ik} \Bigg)' \end{aligned}$$

and the latent random variables $\widetilde{V}_{(uv)ik} = \widetilde{Y}^u_{i1k}(1 - \widetilde{Y}_{i1k})^{1-u}\widetilde{Y}^v_{i2k}(1 - \widetilde{Y}_{i2k})^{1-v}$, for $u, v \in \{0, 1\}$.

If the latent data $\widetilde{\mathbf{Y}}$ were observed, the posterior distributions for the prevalence parameters in $\mathbf{p}$ and the test accuracies in $\boldsymbol{\delta}$ would be fully determined. However, because the true individual statuses in $\widetilde{\mathbf{Y}}$ are not observed, we develop a Gibbs sampler to enable posterior inference. Let $\widetilde{\mathbf{Y}}_{k(i)} = (\widetilde{\mathbf{Y}}'_{1k}, ..., \widetilde{\mathbf{Y}}'_{i-1,k}, \widetilde{\mathbf{Y}}'_{i+1,k}, ..., \widetilde{\mathbf{Y}}'_{c_k k})'$ denote the vector of latent responses in the kth group for all individuals except the ith one. The conditional distribution of $\widetilde{\mathbf{V}}_{ik} = (\widetilde{V}_{(00)ik}, \widetilde{V}_{(10)ik}, \widetilde{V}_{(01)ik}, \widetilde{V}_{(11)ik})'$ given

$\{\widetilde{\mathbf{Y}}_{k(i)}, \mathbf{p}, \boldsymbol{\delta}, \mathbf{Y}, \mathbf{Z}\}$ is multinomial with cell probabilities $\zeta_{00}^{ik}/\zeta^{ik}$, $\zeta_{10}^{ik}/\zeta^{ik}$, $\zeta_{01}^{ik}/\zeta^{ik}$, and $\zeta_{11}^{ik}/\zeta^{ik}$, where

$$\begin{aligned}
\zeta_{00}^{ik} &= p_{00} \prod_{j=1}^{2} \left(S_{e:j}^{Z_{jk}} \overline{S}_{e:j}^{1-Z_{jk}}\right)^{\gamma_{ijk}} \left(S_{p:j}^{1-Z_{jk}} \overline{S}_{p:j}^{Z_{jk}}\right)^{1-\gamma_{ijk}} \left(S_{p:j}^{1-Y_{ijk}} \overline{S}_{p:j}^{Y_{ijk}}\right)^{I(Z_{+k}>0)} \\
\zeta_{10}^{ik} &= p_{10} S_{e:1}^{Z_{1k}} \overline{S}_{e:1}^{1-Z_{1k}} \left(S_{e:2}^{Z_{2k}} \overline{S}_{e:2}^{1-Z_{2k}}\right)^{\gamma_{i2k}} \left(S_{p:2}^{1-Z_{2k}} \overline{S}_{p:2}^{Z_{2k}}\right)^{1-\gamma_{i2k}} \\
&\quad \times \left(S_{e:1}^{Y_{i1k}} \overline{S}_{e:1}^{1-Y_{i1k}} S_{p:2}^{1-Y_{i2k}} \overline{S}_{p:2}^{Y_{i2k}}\right)^{I(Z_{+k}>0)} \\
\zeta_{01}^{ik} &= p_{01} S_{e:2}^{Z_{2k}} \overline{S}_{e:2}^{1-Z_{2k}} \left(S_{e:1}^{Z_{1k}} \overline{S}_{e:1}^{1-Z_{1k}}\right)^{\gamma_{i1k}} \left(S_{p:1}^{1-Z_{1k}} \overline{S}_{p:1}^{Z_{1k}}\right)^{1-\gamma_{i1k}} \\
&\quad \times \left(S_{e:2}^{Y_{i2k}} \overline{S}_{e:2}^{1-Y_{i2k}} S_{p:1}^{1-Y_{i1k}} \overline{S}_{p:1}^{Y_{i1k}}\right)^{I(Z_{+k}>0)} \\
\zeta_{11}^{ik} &= p_{11} \prod_{j=1}^{2} S_{e:j}^{Z_{jk}} \overline{S}_{e:j}^{1-Z_{jk}} \left(S_{e:j}^{Y_{ijk}} \overline{S}_{e:j}^{1-Y_{ijk}}\right)^{I(Z_{+k}>0)},
\end{aligned}$$

$\zeta^{ik} = \sum_{u=0}^{1} \sum_{v=0}^{1} \zeta_{uv}^{ik}$, and $\gamma_{ijk} = I(\sum_{i' \neq i} \widetilde{Y}_{i'jk} > 0)$. Note that by sampling $\widetilde{\mathbf{V}}_{ik}$ from this conditional distribution, $\widetilde{\mathbf{Y}}_{ik} = (\widetilde{V}_{(10)ik} + \widetilde{V}_{(11)ik}, \widetilde{V}_{(01)ik} + \widetilde{V}_{(11)ik})'$; in other words, the true individual statuses in $\widetilde{\mathbf{Y}}_{ik}$ are uniquely determined.

Using the full conditional distributions of $\mathbf{p}$, $\boldsymbol{\delta}$, and $\widetilde{\mathbf{V}}_{ik}$ described above, we now outline our Gibbs sampler to implement a Bayesian analysis with the observed data from the pooling algorithm described in Section 2.2:

GIBBS SAMPLER

1. Initialize $\widetilde{\mathbf{Y}}_{ik}^{(0)} = (\widetilde{Y}_{i1k}^{(0)}, \widetilde{Y}_{i2k}^{(0)})'$, for $i = 1, 2, ..., c_k$ and $k = 1, 2, ..., K$. Set $d = 1$.

2. Sample $\mathbf{p}^{(d)}$ from $\mathbf{p}|\widetilde{\mathbf{Y}}^{(d-1)} \sim \text{Dirichlet}(\boldsymbol{\Psi})$, where $\widetilde{\mathbf{Y}}^{(d-1)}$ is the collection of all $\widetilde{\mathbf{Y}}_{ik}^{(d-1)}$,s.

3. Sample $S_{e:j}^{(d)}$ from $S_{e:j}|\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}^{(d-1)} \sim \text{beta}(a_{S_{e:j}}^*, b_{S_{e:j}}^*)$ and sample $S_{p:j}^{(d)}$ from $S_{p:j}|\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}^{(d-1)} \sim \text{beta}(a_{S_{p:j}}^*, b_{S_{p:j}}^*)$ for $j = 1, 2$. Set $\boldsymbol{\delta}^{(d)} = (S_{e:1}^{(d)}, S_{e:2}^{(d)}, S_{p:1}^{(d)}, S_{p:2}^{(d)})'$.

4. For $i = 1, ..., c_k$ and $k = 1, 2, ..., K$, sample $\widetilde{\mathbf{V}}_{ik}^{(d)} = (\widetilde{V}_{(00)ik}^{(d)}, \widetilde{V}_{(10)ik}^{(d)}, \widetilde{V}_{(01)ik}^{(d)}, \widetilde{V}_{(11)ik}^{(d)})'$ from

$$\widetilde{\mathbf{V}}_{ik}|\widetilde{\mathbf{Y}}_{k(i)}^{(d)}, \mathbf{p}^{(d)}, \boldsymbol{\delta}^{(d)}, \mathbf{Y}, \mathbf{Z} \sim \text{multinomial}\left\{1, (\zeta_{00}^{ik}/\zeta^{ik}, \zeta_{10}^{ik}/\zeta^{ik}, \zeta_{01}^{ik}/\zeta^{ik}, \zeta_{11}^{ik}/\zeta^{ik})'\right\},$$

where $\widetilde{\mathbf{Y}}_{k(i)}^{(d)} = (\widetilde{\mathbf{Y}}_{1k}^{(d)'}, ..., \widetilde{\mathbf{Y}}_{i-1,k}^{(d)'}, \widetilde{\mathbf{Y}}_{i+1,k}^{(d-1)'}, ..., \widetilde{\mathbf{Y}}_{c_k k}^{(d-1)'})'$.

Set $\widetilde{\mathbf{Y}}_{ik}^{(d)} = (\widetilde{V}_{(10)ik}^{(d)} + \widetilde{V}_{(11)ik}^{(d)}, \widetilde{V}_{(01)ik}^{(d)} + \widetilde{V}_{(11)ik}^{(d)})'$.

5. Set $d = d + 1$.

6. Repeat steps 2-5 while $d < G$, the number of Gibbs iterates.

Prior elicitation

We now discuss prior model specification for the assay accuracies $S_{e:j}$ and $S_{p:j}$ and the prevalence parameter $\mathbf{p}$. A noninformative approach might specify flat priors for all parameters; i.e., $S_{e:j} \sim \text{beta}(1,1)$, $S_{p:j} \sim \text{beta}(1,1)$, and $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. In fact, we demonstrate in Section 2.4 that posterior estimates of $\mathbf{p}$ and $\boldsymbol{\delta}$ are close to the true values even when these priors are used. Of course, properly chosen informative prior models should increase posterior precision.

In most screening situations, there will be information readily available for researchers to formulate informative priors for $S_{e:j}$ and $S_{p:j}$. Before a diagnostic assay is introduced for commercial use, its performance is assessed in pilot studies involving known positive and known negative specimens. Data from these studies can be used to elicit sensible beta hyperparameters. For example, the most recent product literature for the Aptima Combo 2 Assay, which is available at `http://www.hologic.com`, summarizes a pilot study describing the assay's performance with urine and swab specimens from females and males. This literature documents that among 127 known NG-positive female urine samples, 116 tested positively with the assay, giving an estimated sensitivity of 0.913. Note that in the absence of additional information, the estimation methods in Tebbs et al. (2013) might require one to essentially treat 0.913 as the "true" sensitivity, disregarding the fact that this is only an estimate. In practice, it is important to acknowledge that assay performance may depend on a variety of factors, including those related to implementation and perhaps even the population

being tested (CDC, 2015). As we show in Appendix A, a beta(117,12) distribution is consistent with the pilot data described above.

To model the prevalence parameter $\mathbf{p}$, we use a class of distributions motivated by the power prior class described in Ibrahim et al. (2015). This class assumes the availability of "historical data" and amalgamates their likelihood function with a prior distribution assumed for them. Applying this idea to our problem, we specify $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0)$ as a prior distribution. The value $a_0 \in [0, 1]$ controls the amount of weight given to the historical data (the so-called "precision parameter"), N_0 is the historical data sample size, and $\mathbf{p}_0$ is an estimate of $\mathbf{p}$ obtained from the historical data. This family of priors is ideal for the large-scale screening applications we consider, where, for example, previous years' testing outcomes can be treated as historical data; see Section 2.5. Note that choosing $a_0 = 0$ gives $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. As a_0 increases, the amount of weight given to the historical data increases.

2.4 Simulation evidence

We performed simulation studies to assess the characteristics of our estimation procedure. We used different values of the prevalence parameter $\mathbf{p}$, different prior models for $S_{e:j}$ and $S_{p:j}$, and prior models for $\mathbf{p}$ which incorporate various levels of historical information. A subset of the results is given herein; complete results are in Appendix A.

Simulation description

We generated individual true statuses from a multinomial distribution with cell probabilities in $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$. We took $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$ to consider moderate-level infections and $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$ to allow for rarer infections. Note that the marginal infection probabilities are 11 and 10 percent for the first case and 4 and 2 percent for the second case. These probabilities are representative of our

CT/NG data application in Section 2.5. In all cases, we used $N = 1000$ individuals. This sample size is much smaller than our application in Section 2.5 but is sufficient to illustrate our main findings.

To simulate the observed diagnoses from the algorithm in Section 2.2, we took the individual true statuses $\widetilde{\mathbf{Y}}_{ik}$ and randomly assigned them to pools of size c_k^*, where c_k^* is the pool size that minimizes the expected number of tests needed to diagnose each individual using the algorithm in Section 2.2 (i.e., it is the most "cost-effective" choice); see Tebbs et al. (2013). Testing responses for pools in Stage 1 were then determined by simulating $Z_{jk} \sim \text{Bernoulli}\{S_{e:j}\widetilde{Z}_{jk} + \overline{S}_{p:j}(1 - \widetilde{Z}_{jk})\}$, where $\widetilde{Z}_{jk} = I(\sum_{i=1}^{c_k^*} \widetilde{Y}_{ijk} > 0)$. In Stage 2, individual testing responses for those pools testing positively in Stage 1 (for either infection) were simulated as $Y_{ijk} \sim \text{Bernoulli}\{S_{e:j}\widetilde{Y}_{ijk} + \overline{S}_{p:j}(1-\widetilde{Y}_{ijk})\}$. Throughout our investigation, we took $S_{e:j} = 0.95$ and $S_{p:j} = 0.99$, for $j = 1, 2$. At each parameter configuration, this entire process was repeated $B = 500$ times.

We used different prior distributions for $\mathbf{p}$ and $\boldsymbol{\delta} = (S_{e:1}, S_{e:2}, S_{p:1}, S_{p:2})'$, including those which incorporated no information (i.e., flat priors) and those which were highly informative. For the prevalence parameter, we took $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0)$, where $N_0 = 1000$ is the size of the historical data set, $\mathbf{p}_0$ is the historical estimate, and $a_0 \in \{0, 0.1, 0.2, ..., 1\}$. For the assay accuracies, we used

- Flat priors: $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$, for $j = 1, 2$
- Informative priors: $S_{e:j} \sim \text{beta}(109.0, 6.7)$ and $S_{p:j} \sim \text{beta}(55.2, 1.6)$, for $j = 1, 2$.

The informative distributions have modes located at the true $S_{e:j} = 0.95$ and $S_{p:j} = 0.99$, 5th percentiles of 0.903 and 0.929, and 95th percentiles of 0.973 and 0.996, respectively. Posterior sampling was done using the Gibbs sampler described in Section 2.3. In the initialization step, we specified the individual true statuses to be the di-

agnosed statuses; i.e., $\widetilde{Y}_{ijk}^{(0)} = I(Z_{i1k} + Z_{i2k} > 0, Y_{ijk} = 1)$, for $j = 1, 2$, and later noted that the performance of our methods was invariant to the choice of initialization. At each configuration, we took $G = 10000$ Gibbs iterates after discarding the first 500. Posterior estimates were calculated from the full set of 10000 Gibbs iterates in each simulation setting (i.e., we did not thin the posterior samples).

Simulation results

We present results for the moderate-level infection case $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$ using flat priors for $S_{e:j}$ and $S_{p:j}$ in Figures 2.1 and 2.2; complete results for all settings described in Section 2.4 are given in Appendix A.

Figure 2.1 provides a summary of the point estimates for $\mathbf{p}$ for all $a_0 \in \{0, 0.1, ..., 1\}$; specifically, the figure shows 5th, 25th, 50th, 75th, and 95th percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$. Figure 2.2 shows the same summary, but for estimates of the assay accuracy parameters in $\boldsymbol{\delta} = (S_{e:1}, S_{e:2}, S_{p:1}, S_{p:2})'$. Recall that when $a_0 = 0$, the prior distribution for $\mathbf{p}$ is Dirichlet($\mathbf{1}_4$); i.e., no historical information is used. Even in this situation with flat priors for $S_{e:j}$ and $S_{p:j}$, median estimates for all prevalence parameters (Figure 2.1) and assay accuracies (Figure 2.2) are on target. In other words, providing no information in the prior distributions still allows for accurate inference on all model parameters. We find this phenomenon to be quite remarkable; not only can one estimate the prevalence parameter $\mathbf{p}$ but one can also accurately estimate the sensitivity and specificity parameters in $\boldsymbol{\delta}$. This desirable feature of our approach is clearly a byproduct of using a single discriminating assay in both stages; retesting individuals in Stage 2−whenever a master pool in Stage 1 tests positively for one infection or both−provides an abundance of information on the assay's accuracy for both infections. In single-infection group testing applications, assay accuracy parameters cannot be identified without retesting or making use of historical data (Johnson and Pearson, 1999; Hanson et al., 2006; Zhang et al., 2014).

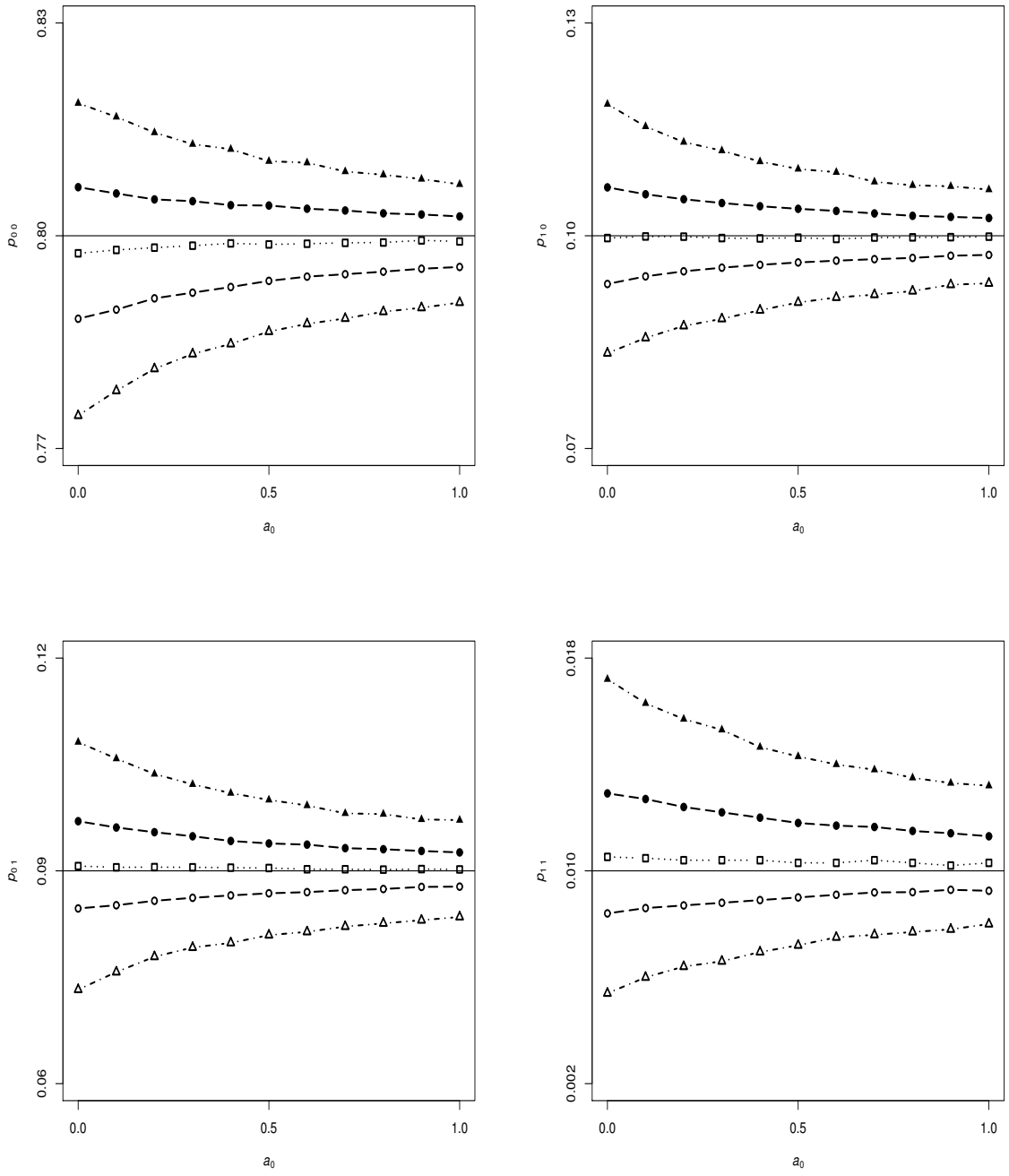

Figure 2.1: Simulation results with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$ are provided. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

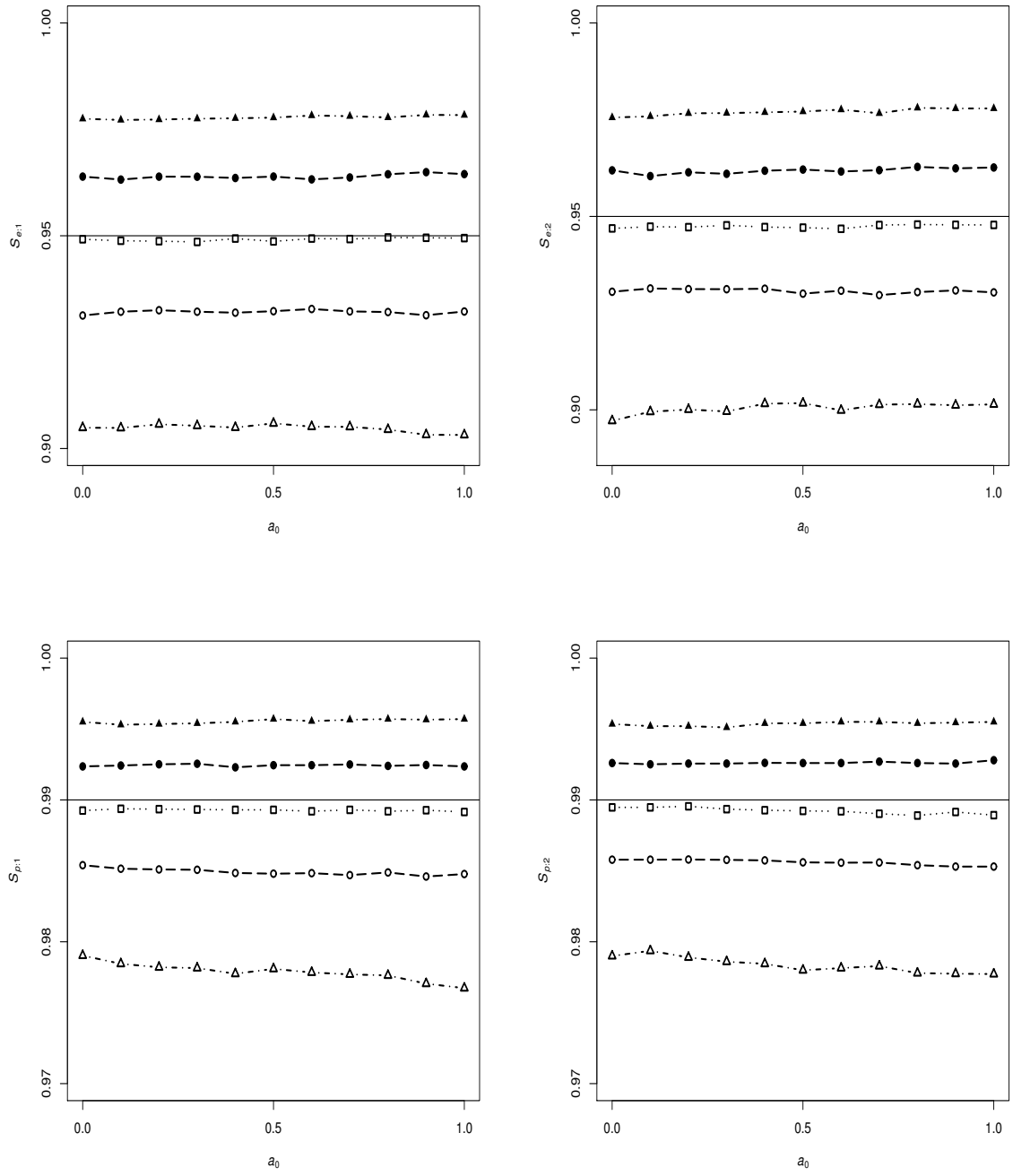

Figure 2.2: Simulation results with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $S_{e:j}$ and $S_{p:j}$ are provided. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

Figures 2.1 and 2.2 also show the impact of adding historical information in the Dirichlet($\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0$) prior distribution for $\mathbf{p}$. For the prevalence parameters in Figure 2.1, posterior distributions tighten noticeably as more information is added (i.e., as a_0 increases), while for the assay accuracies in Figure 2.2, posterior variability is largely unaffected. Figures 2.1 and 2.2 were constructed assuming that the historical estimate $\mathbf{p}_0$ matched the true value of $\mathbf{p}$. However, we obtained nearly identical results when $\mathbf{p}_0$ was misspecified, especially in situations with larger N like our data application in Section 2.5. Of course, the fact that prevalence parameters in $\mathbf{p}$ are estimated well in the absence of prior information ($a_0 = 0$) should comfort the researcher unwilling to place too much faith in a historical prevalence estimate or when no historical data are available. Following the recommendation of an anonymous referee, we also examined how misspecified prior distributions for the assay accuracies $S_{e:j}$ and $S_{p:j}$ could impact our Bayesian estimates and, simultaneously, how using incorrect values of $S_{e:j}$ and $S_{p:j}$ could affect the maximum likelihood estimates in Tebbs et al. (2013). Simulation results summarizing this examination are provided in Appendix A. When applied to $N = 1000$ individuals as before, Bayesian estimates for both $\mathbf{p}$ and $\boldsymbol{\delta}$ remain generally on target even under severe prior model misspecification, whereas the estimates of $\mathbf{p}$ from Tebbs et al. (2013) calculated using incorrect values of $S_{e:j}$ and $S_{p:j}$ can be severely biased.

Simulation results for the other settings; i.e., for (i) moderate-level infections, informative assay priors; (ii) rare infections, flat assay priors; and (iii) rare infections, informative assay priors, are given in Appendix A. Not surprisingly, posterior distributions for assay accuracies tighten when informative priors are used. However, one observation does warrant a remark. When the marginal disease probabilities $p_{10} + p_{11}$ and $p_{01} + p_{11}$ are both small, fewer pools will be resolved in Stage 2 and the sensitivity parameters can be underestimated. Naturally, fewer retests will curtail the information needed to estimate both $S_{e:j}$ and $S_{p:j}$ accurately, but this will have a

more pronounced effect on assay sensitivity because fewer individuals will be truly positive. At the same time, it should be noted that this "underestimation behavior" disappeared completely when we increased the sample size in our simulation to $N = 10000$ individuals (results not shown), a sample size more in line with our data application in Section 2.5.

2.5 Infertility Prevention Project data

To illustrate our methodology with real data, we use CT/NG testing outcomes collected by the Nebraska Public Health Laboratory (NPHL) during 2008 and 2009. As part of their IPP screening program, there were 23,146 individuals tested in 2008 and 27,551 individuals tested in 2009. Our data set consists of individual diagnoses for each infection for both 2008 and 2009. We implement our Bayesian group testing estimation methods for the 2009 individuals, using the 2008 results as historical data.

To perform the 2009 analysis, we first cross-classified individuals in both years according to their gender and specimen type (swab or urine). This was done because operating characteristics for assays commonly used for CT/NG detection depend on these factors (Gaydos et al., 2010). Within each of the four gender/specimen type strata, we analyzed the 2008 individual testing results to formulate a historical estimate $\mathbf{p}_0$ that accounts for potential misclassification. These historical estimates are shown in Table 2.1; for more details on how these were computed, see Appendix A. Also given in Table 2.1 are informative beta prior distributions for $S_{e:j}$ and $S_{p:j}$ ($j = 1$ for CT; $j = 2$ for NG), which, for purposes of illustration, were chosen to be consistent with the pilot studies presented in the Aptima Combo 2 Assay product literature. We use Appendix A to provide precise details on how these informative prior distributions were chosen. Finally, for the prevalence parameter $\mathbf{p}$ in each stratum, we took $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4 + a_0 N_0 \mathbf{p}_0)$, where N_0 is the 2008 stratum size, and examined the $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ cases separately. Note that taking $a_0 = 0$

ignores the 2008 information about the prevalence whereas the $a_0 = 1$ case weighs the 2008 information heavily. One might view taking $a_0 = 0.5$ as a compromise between the two extremes.

To emulate the use of group testing with the 2009 individuals, we assigned them to pools of size c_k^* which was determined using the 2008 historical information in Table 2.1. This was done within each gender/specimen type stratum and chronologically based on the individual specimen's arrival date at the NPHL. Using the 2009 individual testing results, acknowledging that these are potentially incorrect, we first simulated the individual true responses $\widetilde{\mathbf{Y}}_{ik} = (\widetilde{Y}_{i1k}, \widetilde{Y}_{i2k})'$; a description of how this was done is given in Appendix A. Observed diagnoses from the pooling algorithm in Section 2.2 were then simulated using $Z_{jk} \sim \text{Bernoulli}\{S_{e:j}\widetilde{Z}_{jk} + \overline{S}_{p:j}(1 - \widetilde{Z}_{jk})\}$ and $Y_{ijk} \sim \text{Bernoulli}\{S_{e:j}\widetilde{Y}_{ijk} + \overline{S}_{p:j}(1 - \widetilde{Y}_{ijk})\}$ using the Aptima Combo 2 Assay pilot study point estimates for $S_{e:j}$ and $S_{p:j}$. To average over the effect of simulation error, we repeated this exercise $B = 500$ times, leaving us with 500 simulated group testing data sets for each gender/specimen type stratum in 2009.

For each stratum and for each choice of $a_0 \in \{0, 0.5, 1\}$, Table 2.2 shows the 2009 estimation results for the population-level parameters

$$
\begin{aligned}
p_{00} &= \text{proportion of individuals negative for both CT and NG} \\
p_{10} &= \text{proportion of individuals positive for CT, but negative for NG} \\
p_{01} &= \text{proportion of individuals negative for CT, but positive for NG} \\
p_{11} &= \text{proportion of individuals positive for both CT and NG.}
\end{aligned}
$$

Table 2.1: Nebraska 2008 historical information for CT/NG. The historical estimate $\mathbf{p}_0 = (p_{00(0)}, p_{10(0)}, p_{01(0)}, p_{11(0)})'$ was calculated using the 2008 individual diagnoses (accounting for possible misclassification; see Appendix A). Stratum sample sizes N_0 are given. Priors for $S_{e:j}$ and $S_{p:j}$ were determined using pilot data from the Aptima Combo 2 Assay product literature (see Appendix A). The master pool size c_k^* minimizes the expected number of tests per individual as described in Tebbs et al. (2013).

Stratum	CT	NG	$\mathbf{p}_0$	Prior for S_e	Prior for S_p	Pool size
Male/Urine $N_0 = 3541$	$-$	$-$	$p_{00(0)} = 0.929$	CT: beta$(277, 7)$ NG: beta$(325, 6)$	CT: beta$(802, 13)$ NG: beta$(803, 4)$	$c_k^* = 4$
	$+$	$-$	$p_{10(0)} = 0.061$			
	$-$	$+$	$p_{01(0)} = 0.005$			
	$+$	$+$	$p_{11(0)} = 0.004$			
Male/Swab $N_0 = 2826$	$-$	$-$	$p_{00(0)} = 0.848$	CT: beta$(261, 12)$ NG: beta$(320, 4)$	CT: beta$(775, 21)$ NG: beta$(765, 18)$	$c_k^* = 3$
	$+$	$-$	$p_{10(0)} = 0.103$			
	$-$	$+$	$p_{01(0)} = 0.032$			
	$+$	$+$	$p_{11(0)} = 0.016$			
Female/Urine $N_0 = 2338$	$-$	$-$	$p_{00(0)} = 0.907$	CT: beta$(198, 12)$ NG: beta$(117, 12)$	CT: beta$(1171, 14)$ NG: beta$(1348, 11)$	$c_k^* = 4$
	$+$	$-$	$p_{10(0)} = 0.074$			
	$-$	$+$	$p_{01(0)} = 0.006$			
	$+$	$+$	$p_{11(0)} = 0.013$			
Female/Swab $N_0 = 14441$	$-$	$-$	$p_{00(0)} = 0.948$	CT: beta$(196, 13)$ NG: beta$(127, 2)$	CT: beta$(1155, 29)$ NG: beta$(1336, 18)$	$c_k^* = 5$
	$+$	$-$	$p_{10(0)} = 0.047$			
	$-$	$+$	$p_{01(0)} = 0.001$			
	$+$	$+$	$p_{11(0)} = 0.005$			

Table 2.2: Nebraska CT/NG prevalence estimation results for 2009. Bayesian estimates (Bayes) are posterior medians averaged over $B = 500$ data sets; BSE is the average of the standard deviations calculated from posterior samples of the $B = 500$ data sets. Values of $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ are used to incorporate different amounts of historical information for $\mathbf{p}$ as described in Section 2.5. Prior distributions for $S_{e:j}$ and $S_{p:j}$, where $j = 1$ for CT and $j = 2$ for NG, are given in Table 2.1. Maximum likelihood estimates, calculated from Tebbs et al. (2013), are averaged over the same 500 data sets; the entries under SE are the averaged standard errors. Stratum sample sizes N are given.

			Maximum likelihood		Bayes ($a_0 = 0$)		Bayes ($a_0 = 0.5$)		Bayes ($a_0 = 1$)	
Stratum	CT	NG	Estimate	SE	Estimate	BSE	Estimate	BSE	Estimate	BSE
Male/Urine $N = 6139$	$-$	$-$	$\widehat{p}_{00} = 0.924$	0.0035	$\widehat{p}_{00} = 0.923$	0.0037	$\widehat{p}_{00} = 0.926$	0.0029	$\widehat{p}_{00} = 0.927$	0.0025
	$+$	$-$	$\widehat{p}_{10} = 0.061$	0.0032	$\widehat{p}_{10} = 0.061$	0.0034	$\widehat{p}_{10} = 0.061$	0.0027	$\widehat{p}_{10} = 0.061$	0.0023
	$-$	$+$	$\widehat{p}_{01} = 0.008$	0.0012	$\widehat{p}_{01} = 0.008$	0.0012	$\widehat{p}_{01} = 0.007$	0.0009	$\widehat{p}_{01} = 0.006$	0.0007
	$+$	$+$	$\widehat{p}_{11} = 0.007$	0.0011	$\widehat{p}_{11} = 0.007$	0.0011	$\widehat{p}_{11} = 0.006$	0.0008	$\widehat{p}_{11} = 0.006$	0.0007
Male/Swab $N = 1910$	$-$	$-$	$\widehat{p}_{00} = 0.831$	0.0091	$\widehat{p}_{00} = 0.831$	0.0096	$\widehat{p}_{00} = 0.837$	0.0074	$\widehat{p}_{00} = 0.841$	0.0062
	$+$	$-$	$\widehat{p}_{10} = 0.119$	0.0079	$\widehat{p}_{10} = 0.119$	0.0085	$\widehat{p}_{10} = 0.113$	0.0064	$\widehat{p}_{10} = 0.111$	0.0054
	$-$	$+$	$\widehat{p}_{01} = 0.034$	0.0043	$\widehat{p}_{01} = 0.034$	0.0044	$\widehat{p}_{01} = 0.033$	0.0035	$\widehat{p}_{01} = 0.033$	0.0030
	$+$	$+$	$\widehat{p}_{11} = 0.015$	0.0029	$\widehat{p}_{11} = 0.015$	0.0030	$\widehat{p}_{11} = 0.016$	0.0024	$\widehat{p}_{11} = 0.016$	0.0021
Female/Urine $N = 4972$	$-$	$-$	$\widehat{p}_{00} = 0.920$	0.0041	$\widehat{p}_{00} = 0.919$	0.0044	$\widehat{p}_{00} = 0.915$	0.0035	$\widehat{p}_{00} = 0.913$	0.0030
	$+$	$-$	$\widehat{p}_{10} = 0.066$	0.0038	$\widehat{p}_{10} = 0.067$	0.0041	$\widehat{p}_{10} = 0.069$	0.0032	$\widehat{p}_{10} = 0.071$	0.0028
	$-$	$+$	$\widehat{p}_{01} = 0.004$	0.0010	$\widehat{p}_{01} = 0.004$	0.0011	$\widehat{p}_{01} = 0.005$	0.0009	$\widehat{p}_{01} = 0.005$	0.0008
	$+$	$+$	$\widehat{p}_{11} = 0.009$	0.0014	$\widehat{p}_{11} = 0.009$	0.0015	$\widehat{p}_{11} = 0.011$	0.0012	$\widehat{p}_{11} = 0.011$	0.0011
Female/Swab $N = 14530$	$-$	$-$	$\widehat{p}_{00} = 0.949$	0.0019	$\widehat{p}_{00} = 0.949$	0.0023	$\widehat{p}_{00} = 0.949$	0.0017	$\widehat{p}_{00} = 0.948$	0.0015
	$+$	$-$	$\widehat{p}_{10} = 0.045$	0.0019	$\widehat{p}_{10} = 0.045$	0.0023	$\widehat{p}_{10} = 0.046$	0.0017	$\widehat{p}_{10} = 0.047$	0.0014
	$-$	$+$	$\widehat{p}_{01} = 0.001$	0.0001	$\widehat{p}_{01} = 0.001$	0.0002	$\widehat{p}_{01} = 0.001$	0.0001	$\widehat{p}_{01} = 0.001$	0.0001
	$+$	$+$	$\widehat{p}_{11} = 0.005$	0.0006	$\widehat{p}_{11} = 0.005$	0.0006	$\widehat{p}_{11} = 0.005$	0.0005	$\widehat{p}_{11} = 0.005$	0.0004

Table 2.3: Bayesian assay accuracy estimates from 2009. Bayesian estimates (Bayes) are posterior medians averaged over $B = 500$ data sets; BSE is the average of the standard deviations calculated from posterior samples of the $B = 500$ data sets. Values of $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ are used to incorporate different amounts of historical information for $\mathbf{p}$ as described in Section 2.5. Prior distributions for $S_{e:j}$ and $S_{p:j}$, where $j = 1$ for CT and $j = 2$ for NG, are given in Table 2.1. Stratum sample sizes N are given.

		Bayes ($a_0 = 0$)		Bayes ($a_0 = 0.5$)		Bayes ($a_0 = 1$)	
Stratum	Accuracy	Estimate	BSE	Estimate	BSE	Estimate	BSE
	$S_{e:1} = 0.979$	$\widehat{S}_{e:1} = 0.977$	0.0076	$\widehat{S}_{e:1} = 0.978$	0.0074	$\widehat{S}_{e:1} = 0.978$	0.0073
Male/Urine	$S_{e:2} = 0.985$	$\widehat{S}_{e:2} = 0.983$	0.0069	$\widehat{S}_{e:2} = 0.984$	0.0066	$\widehat{S}_{e:2} = 0.984$	0.0065
$N = 6139$	$S_{p:1} = 0.985$	$\widehat{S}_{p:1} = 0.985$	0.0030	$\widehat{S}_{p:1} = 0.985$	0.0030	$\widehat{S}_{p:1} = 0.984$	0.0030
	$S_{p:2} = 0.996$	$\widehat{S}_{p:2} = 0.996$	0.0012	$\widehat{S}_{p:2} = 0.996$	0.0012	$\widehat{S}_{p:2} = 0.996$	0.0012
	$S_{e:1} = 0.959$	$\widehat{S}_{e:1} = 0.958$	0.0107	$\widehat{S}_{e:1} = 0.960$	0.0102	$\widehat{S}_{e:1} = 0.961$	0.0099
Male/Swab	$S_{e:2} = 0.991$	$\widehat{S}_{e:2} = 0.989$	0.0061	$\widehat{S}_{e:2} = 0.989$	0.0060	$\widehat{S}_{e:2} = 0.989$	0.0060
$N = 1910$	$S_{p:1} = 0.975$	$\widehat{S}_{p:1} = 0.974$	0.0048	$\widehat{S}_{p:1} = 0.973$	0.0049	$\widehat{S}_{p:1} = 0.973$	0.0049
	$S_{p:2} = 0.978$	$\widehat{S}_{p:2} = 0.979$	0.0034	$\widehat{S}_{p:2} = 0.979$	0.0035	$\widehat{S}_{p:2} = 0.979$	0.0035
	$S_{e:1} = 0.947$	$\widehat{S}_{e:1} = 0.946$	0.0115	$\widehat{S}_{e:1} = 0.944$	0.0114	$\widehat{S}_{e:1} = 0.942$	0.0115
Female/Urine	$S_{e:2} = 0.913$	$\widehat{S}_{e:2} = 0.909$	0.0217	$\widehat{S}_{e:2} = 0.905$	0.0220	$\widehat{S}_{e:2} = 0.904$	0.0222
$N = 4972$	$S_{p:1} = 0.989$	$\widehat{S}_{p:1} = 0.989$	0.0026	$\widehat{S}_{p:1} = 0.989$	0.0025	$\widehat{S}_{p:1} = 0.989$	0.0025
	$S_{p:2} = 0.993$	$\widehat{S}_{p:2} = 0.993$	0.0016	$\widehat{S}_{p:2} = 0.993$	0.0015	$\widehat{S}_{p:2} = 0.993$	0.0015
	$S_{e:1} = 0.942$	$\widehat{S}_{e:1} = 0.941$	0.0112	$\widehat{S}_{e:1} = 0.939$	0.0107	$\widehat{S}_{e:1} = 0.938$	0.0106
Female/Swab	$S_{e:2} = 0.992$	$\widehat{S}_{e:2} = 0.986$	0.0104	$\widehat{S}_{e:2} = 0.987$	0.0098	$\widehat{S}_{e:2} = 0.987$	0.0097
$N = 14530$	$S_{p:1} = 0.976$	$\widehat{S}_{p:1} = 0.976$	0.0029	$\widehat{S}_{p:1} = 0.976$	0.0028	$\widehat{S}_{p:1} = 0.976$	0.0028
	$S_{p:2} = 0.987$	$\widehat{S}_{p:2} = 0.987$	0.0013	$\widehat{S}_{p:2} = 0.987$	0.0014	$\widehat{S}_{p:2} = 0.987$	0.0013

We used our estimation approach in Section 2.3 to calculate posterior medians for each of the 500 data sets; the results shown in Table 2.2 are averages across the data sets. For comparison purposes, we also show the maximum likelihood (ML) results averaged over the same 500 data sets. Table 2.3 shows the 2009 (Bayesian) estimation results for $\boldsymbol{\delta} = (S_{e:1}, S_{e:2}, S_{p:1}, S_{p:2})'$ under the same settings. Recall that for each data set, our Bayesian approach estimates $\mathbf{p}$ and $\boldsymbol{\delta}$ simultaneously whereas the ML approach from Tebbs et al. (2013) regards $\boldsymbol{\delta}$ as fixed and known.

Table 2.2 shows that the averaged ML estimates and the averaged Bayesian posterior median estimates are very similar in most settings. For the male/swab and female/urine cohorts, the average posterior median estimates are slightly different than the average ML estimates when the 2008 historical information (Table 2.1) is weighed more heavily; i.e., the posterior estimates are more strongly attracted to the historical estimates when a_0 increases. The column in Table 2.2 labeled "SE" is the averaged standard error of the 500 ML estimates, and the column labeled "BSE" is the analogous Bayesian measure of variability which we calculated as follows. For each data set, the entire Gibbs chain included 3500 iterates; the first 500 were discarded (as in Section 2.4) and the sample standard deviation was calculated using every 6th iterate from the last 3000. Thinning was used to remove autocorrelation from the successive iterates and left us with a within-data set measure of variation based on 500 posterior draws. BSE was then calculated as the average of these standard deviations across the $B = 500$ data sets, a measure that can be compared with SE.

One of the surprising discoveries from Tebbs et al. (2013) was that, despite the reduction in the number of tests needed, ML estimates of $\mathbf{p}$ from using group testing were more efficient than those from individual testing; see Liu et al. (2012) for discussion on this same phenomenon in single disease settings. Comparing the SE and BSE columns in Table 2.2, one notes that further efficiency gains are possible

using our Bayesian approach when historical information on the prevalence parameter is utilized (i.e., when $a_0 = 0.5$ and when $a_0 = 1$). When the 2008 information is ignored ($a_0 = 0$), it is not surprising that the Bayesian estimates are less precise than the ML estimates. However, they are nearly as precise and this is despite the fact that the assay accuracy parameters in $\boldsymbol{\delta}$ are estimated simultaneously along with the prevalence parameters in $\mathbf{p}$ (see Table 2.3). One might conjecture that the informative priors we used for $\boldsymbol{\delta}$ in Table 2.1 may be partly responsible for the increased efficiency in estimation when historical information about $\mathbf{p}$ is incorporated. However, we have repeated this same analysis using flat priors for $S_{e:j}$ and $S_{p:j}$ and have discovered largely the same findings. The tables summarizing this analysis are shown in Appendix A.

2.6 Discussion

We have presented a Bayesian approach to estimate the prevalence of multiple diseases from group testing data, motivated by nationwide CT/NG screening activities to identify infected individuals and to estimate the prevalence of each infection. Our methods allow researchers to incorporate uncertainty in diagnostic assays and also information from previous periods of screening. When compared to the estimation methods in Tebbs et al. (2013), our approach is more flexible; one can obtain estimates of the prevalence parameters while simultaneously estimating assay accuracy. Furthermore, our Bayesian estimates are nearly as precise as ML estimates when no historical information is provided and are potentially more precise otherwise. The web site `www.chrisbilder.com/grouptesting` contains R programs that implement our estimation techniques for the algorithm in Section 2.2. Our programs determine appropriate beta prior distributions for $\boldsymbol{\delta}$ based on pilot studies like those described in Section 2.3. Appendix A shows explicitly how these prior distributions can be constructed.

Our augmented likelihood function $\pi(\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}|\boldsymbol{\theta})$ in Equation (2.1) is constructed by making two simplifying assumptions. First, we assume that all testing outcomes associated with the same master pool (including any individual retest responses from it) are conditionally independent given the true individual statuses in the pool. This assumption is pervasive in the group testing literature for single infections and may be reasonable when misclassification is driven primarily by factors related to test implementation. Second, at both the master pool (Stage 1) and individual testing (Stage 2) levels, we assume that $S_{e:j}$ and $S_{p:j}$ for one infection do not depend on the true status of the other infection. Future research in group testing could investigate ways to avoid making either or both assumptions. For single infections, Wang et al. (2015) relax the conditional independence assumption by relating binary testing outcomes to latent biomarker levels. Relaxing the second assumption with multiple infections might be possible by borrowing ideas from the recent causal inference literature (Hudgens and Halloran, 2008).

Although we have focused on screening for CT/NG, the group testing algorithm in Section 2.2 could be implemented in other situations when a discriminating assay is used to test for two or more infections. For example, the CDC has recently proposed that a discriminating assay for HIV-1 and HIV-2 replace the more traditional Western Blot-type assay to improve the detection of acute HIV infections (Branson and Mermin, 2011; Krajden et al., 2014). Furthermore, discriminating multiplex assays, such as the Procleix Ultrio Plus Assay (Hologic/Gen-Probe, Inc., San Diego), are currently available to simultaneously detect HBV, HCV, and HIV in pooled and individual specimens. Generalizing our techniques to estimate the joint prevalence of more than two infections using the algorithm in Section 2.2 is straightforward; see Appendix A.

Our work has assumed that a discriminating multiplex assay is available and that it can be applied to both pools of specimens and individual specimens. In other situations, especially those involving screening donated blood for HBV, HCV, and HIV (Stramer et al., 2013; O'Brien et al., 2012; Schmidt et al., 2010; Stramer et al., 2011), a non-discriminating multiplex assay (i.e., an assay whose result indicates the presence of at least one infection or the absence of all infections) is typically used to test pools in the first stage. Positive pools are resolved in the next stage by applying the same multiplex assay to individuals (to determine which individuals are positive for at least one infection) and then confirmatory, infection-specific assays to positive individuals to diagnose the presence/absence of each infection. In our review of the infectious disease testing literature, we have found that there are numerous testing algorithms currently used with multiplex assays−discriminating and not−for two or more infections. An advantage of the Bayesian framework outlined in Section 2.3 is that it can be suitably adapted to handle different algorithms with little added difficulty.

Chapter 3

Group testing regression models with dilution submodels

Summary: Group testing, where subjects are tested initially in pools, is commonly used to screen individuals for infectious diseases. When testing is performed for a rare disease, group testing may increase the number of false negative test results due to pooled dilution effects. If testing results are used to estimate individual-level disease probabilities in a regression context, estimates of the regression parameters can be severely biased when the dilution effect is ignored. Most existing regression approaches for pooled binary responses assume that assay sensitivity is a known constant and is independent of the pool size. We propose a new regression method that adjusts for a potential dilution effect. We augment existing regression models that have been proposed in the group testing literature with a secondary parametric dilution model for pooled-level sensitivity and estimate all parameters using maximum likelihood. Our approach provides reliable inference for regression parameters in the presence of dilution. Furthermore, we propose a formal hypothesis test that detects dilution in group testing. We illustrate our method using HBV data collected from a prison population in Ireland.

3.1 Introduction

Group testing involves testing pooled biospecimens (e.g., blood, urine, swab, etc.) for a binary characteristic, such as the presence or absence of a sexually transmitted disease. Dorfman (1943) introduced the idea of group testing to screen United States soldiers during World War II and demonstrated that its use can be far more efficient than individual testing. Since then, group testing has been recommended for various large-scale screening applications, for example, to detect HIV (Pilcher et al., 2005), chlamydia and gonorrhea (Lewis et al., 2012), HBV and HCV (Hourfar et al., 2008; Stramer et al., 2013), and herpes simplex virus (Hanson et al., 2007). Group testing is also used in other applications including animal testing (Dhand et al., 2010), genetics (Chi et al., 2009), pollution detection (Wahed et al., 2006), food safety (Fahey et al., 2006), and drug discovery (Remlinger et al., 2006).

Statistical research in group testing can be categorized into two primary areas: classification and estimation. While the goal of classification research is to develop screening methods that reduce the number of tests needed (Kim et al., 2007; McMahan et al., 2012a), the latter aims to estimate either an overall prevalence for a homogeneous population or individual-level probabilities using covariates in a heterogeneous population. If the disease prevalence is rare, pooled testing results provide a sufficient amount of information for estimation without sacrificing large amounts of efficiency. If additional retest results are available, estimation from using group testing can actually be more efficient than individual testing (Liu et al., 2012; Tebbs et al., 2013; Zhang et al., 2013). In recent years, group testing research has explored regression problems where the goal is to estimate individual-level disease probability. For example, Vansteelandt et al. (2000) and Xie (2001) have proposed parametric methods, Wang et al. (2014b) have adopted a semiparametric approach, and Delaigle and Meister (2011) have presented nonparametric regression techniques.

A concern associated with group testing is the so-called "dilution effect." If tests are performed on large pools, positive individuals can be diluted by negative ones, leading to false negative results. Failure to acknowledge such errors may seriously compromise inference in disease screening (Wein and Zenios, 1996) and in estimation (Hung and Swallow, 1999; McMahan et al., 2013). Previous research in the context of a homogeneous population has addressed this issue. Hwang (1976) studied the screening algorithm of Dorfman (1943) in the presence of dilution. Zenios and Wein (1998) proposed hierarchical models for HIV that make use of continuous biomarker responses and (latent) antibody biomarker concentrations. Hung and Swallow (1999) studied group testing robustness to estimate a population proportion. When individual covariate information is available, McMahan et al. (2013), Wang et al. (2015), and Delaigle and Hall (2015) proposed regression methods under the tenuous assumptions that (i) continuous biomarker information is available, and (ii) disease concentration (e.g., optical density readings for HIV) for a pool is the average of the individuals' disease concentrations within the pool. These assumptions might limit the usefulness of these methods in seroprevalence studies where, typically, knowledge of an underlying biomarker distribution is absent.

In this paper, we take a different approach to account for dilution in a group testing regression setting. We specify a parametric function (submodel) for pool-specific sensitivity, similar to the approach taken by Hung and Swallow (1999) for a homogeneous population. Our approach offers two attractive features. First, it does not require information about underlying continuous biomarker distributions. One can easily construct a dilution submodel using any cumulative distribution function even in the absence of information on an assay's pool testing sensitivity. Second, using the information in the regression model covariates, one can within our framework actually perform a hypothesis test to detect dilution. Furthermore, our method estimates pooled-level sensitivity along with the underlying regression function for

disease positivity. To illustrate our method, we consider two commonly used data collection methods for estimation: (i) master pool testing and (ii) Dorfman (1943) decoding. Master pool testing uses test results from the initial master pools only. Dorfman decoding incorporates additional retest results from those pools that test positively.

This paper is organized as follows. In Section 3.2, we describe our model formulation and maximum likelihood estimation techniques. In Section 3.3, we develop a hypothesis test to detect dilution. We perform an extensive simulation study in Section 3.4. We analyze HBV data collected from Irish prisoners in Section 3.5. Finally, we briefly summarize our findings and discuss future research ideas in Section 3.6. Supplementary materials are provided in Appendix B.

3.2 Estimation

A general methodology is proposed for regression analysis of data observed from master pool testing and Dorfman decoding. The initial stage for both of these algorithms involves assigning each of the N individuals to exactly one of J master pools. Let $\widetilde{Y}_{ij} = 1$ if the ith individual assigned to the jth pool is truly positive, $\widetilde{Y}_{ij} = 0$ otherwise, for $i = 1, 2, ..., c_j$ and $j = 1, 2, ..., J$. Herein it is assumed that the $\widetilde{Y}_{ij}$'s are independent random variables with

$$\text{pr}(\widetilde{Y}_{ij} = 1) = g^{-1}(\mathbf{x}_{ij}'\boldsymbol{\beta}) = p_{ij},$$

where $\mathbf{x}_{ij} = (1, x_{ij1}, ..., x_{ijr})'$ is an $(r+1) \times 1$ vector of covariates, $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_r)'$ is the corresponding vector of regression coefficients, and $g(\cdot)$ is a monotone and differentiable link function. Let $\widetilde{Z}_j = 1$ if the jth pool is comprised of at least one infected individual, $\widetilde{Z}_j = 0$ otherwise; i.e., $\widetilde{Z}_j = I(\sum_{i=1}^{c_j} \widetilde{Y}_{ij} > 0)$. Consequently, the $\widetilde{Z}_j$'s are independent random variables with

$$\text{pr}(\widetilde{Z}_j = 1) = 1 - \prod_{i=1}^{c_j}(1 - p_{ij}) = 1 - \prod_{i=1}^{c_j}\{1 - g^{-1}(\mathbf{x}_{ij}'\boldsymbol{\beta})\}.$$

For notational convenience, let $\widetilde{\mathbf{Y}}$, $\widetilde{\mathbf{Z}}$, and $\mathbf{X}$ denote the aggregated collection of individual true statuses, the true statuses of the pools, and the individual level covariate information, respectively. In the presence of imperfect testing, the true statuses of the individuals and pools are unobserved for both master pool testing and Dorfman testing; i.e., the $\widetilde{Y}_{ij}$'s and $\widetilde{Z}_j$'s are latent random variables.

Let Z_j denote the testing for the jth master pool; i.e., $Z_j = 1$ if the jth pool tests positively, $Z_j = 0$ otherwise. The data observed from implementing master pool testing consists of $\mathcal{D}_M = \{\mathbf{Z}, \mathbf{X}\}$, where $\mathbf{Z} = \{Z_1, Z_2..., Z_J\}$. Dorfman decoding resolves positive master pools by retesting each contributing individual separately. Let Y_{ij} denote the testing response observed for the ith individual in the jth pool, if this individual is tested separately, where $Y_{ij} = 1$ if the individual tested positively, $Y_{ij} = 0$ otherwise. The data observed from completing Dorfman decoding is given by $\mathcal{D}_D = \{\mathbf{Z}, \mathbf{Y}, \mathbf{X}\}$, where $\mathbf{Y}$ denotes the collection of all individual testing responses.

To perform the regression analysis, one is left to relate the observed testing data to the individuals' true latent statuses. To accomplish this task, previously proposed regression techniques (e.g., Vansteelandt et al., 2000; Wang et al., 2014) proceed under the assumption that the sensitivity of the assay S_e is a known constant, which does not depend on the size of the pool; i.e., $S_e = \text{pr}(Z_j = 1|\widetilde{Z}_j = 1) = \text{pr}(Z_{j'} = 1|\widetilde{Z}_{j'} = 1)$, for all j and j', even if $c_j \neq c_{j'}$. This assumption also implies $\text{pr}(Y_{ij} = 1|\widetilde{Y}_{ij} = 1) = S_e$ for individual testing. From Bayes' theorem and the Law of Total Probability, one may express the probability that a pool tests positively, given it is truly positive, as

$$\begin{aligned} S_e &= \text{pr}(Z_j = 1|\widetilde{Z}_j = 1) \\ &= \{\text{pr}(\widetilde{Z}_j = 1)\}^{-1} \sum_{k=1}^{c_j} \text{pr}\left(Z_j = 1 \middle| \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right) \text{pr}\left(\sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right), \end{aligned}$$

for $j = 1, 2, ..., J$. In general, for the constant sensitivity assumption to be valid one would have to have $S_e = \text{pr}(Z_j = 1|\sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k)$, for all $k = 1, 2, ..., c_j$; i.e., a pool consisting of one truly positive individual would have the same probability of testing positively as a pool comprised of all positive individuals. In low prevalence

settings, the probability a pool consists of more than one positive is negligible, thus by recalibrating assay thresholds and/or dilution ratios such that $S_e = \text{pr}(Z_j = 1 | \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = 1)$, for all j, could provide a setting in which this traditional assumption is reasonable. In contrast, after considering how diagnostic assays render diagnoses, it would typically be more reasonable to assume that

$$\text{pr}\left(Z_j = 1 \middle| \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right) \leq \text{pr}\left(Z_j = 1 \middle| \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k'\right), \text{ for all } k < k', \quad (3.1)$$

$$\text{pr}\left(Z_j = 1 \middle| \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right) \leq \text{pr}\left(Z_{j'} = 1 \middle| \sum_{i=1}^{c_{j'}} \widetilde{Y}_{ij'} = k\right), \text{ for all } c_j > c_{j'}. \quad (3.2)$$

These two characteristics embody what is commonly referred to as the "dilution effect," and herein a general modeling technique is proposed that accounts for both (3.1) and (3.2).

Dilution submodel

To develop submodels which accurately account for the dilution effect, three primary assumptions are made. First, the assay's sensitivity for individual level testing is known and henceforth is denoted by S_e; i.e., $\text{pr}(Y_{ij} = 1 | \widetilde{Y}_{ij} = 1) = S_e$. Second, a pool comprised of all positives will be diagnosed as such with probability S_e; i.e., $S_e = \text{pr}(Z_j = 1 | \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = c_j)$ for all j. Lastly, the probability a pool tests positively is monotonically increasing in the number of positive individuals assigned to it. These general assumptions emit a large class of candidate models which can account for a dilution effect. Although when one couples the monotonicity assumption with the fact that a probability is being modeled, it is natural to consider a cumulative distribution function as a basis for the development of a dilution submodel.

Parametrically modeling the dilution effect is tantamount to modeling the probability that a pool tests positively given that it contains k infected individuals; i.e., specifying

$$h(k, c_j, \lambda) \;=\; \text{pr}\left(Z_j = 1 \middle| \sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right),$$

where the form of h is known up to the parameter λ, and also h is monotone and differentiable with respect to λ. Through estimating the unknown parameter λ, these models can account for a dramatic (minor) dilution effect, corresponding to λ being large (small), or even no effect when $\lambda = 0$. See Section 3.4 and Appendix B.4 for functions that meet these requirements.

Master pool testing

The data available for modeling when master pool testing is implemented consists of $\mathcal{D}_M = \{\mathbf{Z}, \mathbf{X}\}$. In order to relate the observed testing data to the individuals' latent statuses, it is assumed that the assay's specificity, denoted as S_p, is a known fixed constant which does not depend on the size of the pool; i.e., $S_p = \text{pr}(Z_j = 0 | \widetilde{Z}_j = 0)$ for all j. In practice, after recalibrating assay thresholds and/or diltuion ratios it should be reasonable to proceed under this assumption, because the event $\{\widetilde{Z}_j = 0\}$ can only occur if all contributing individuals are truly negative. We continue to make this assumption for S_p throughout the paper. Using a dilution submodel h as defined in Section 3.2, it follows that

$$\text{pr}(Z_j = 1) = p_j = (1 - S_p) \prod_{i=1}^{c_j} (1 - p_{ij}) + \sum_{k=1}^{c_j} h(k, c_j, \lambda) \, \text{pr}\left(\sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right),$$

where the probability $\text{pr}(\sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k)$ involves the sum of independent but non-identically distributed Bernoulli random variables. This probability can be calculated using the approach outlined in Wang (1993).

The observed data likelihood, based on the responses observed from master pool testing and individual level covariates, is given by

$$L(\boldsymbol{\theta}|\mathcal{D}_M) = \prod_{j=1}^{J} p_j^{Z_j}(1-p_j)^{1-Z_j}, \tag{3.3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \lambda)'$. The maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, denoted as $\widehat{\boldsymbol{\theta}}$, is obtained by maximizing $L(\boldsymbol{\theta}|\mathcal{D}_M)$. To conduct large-sample inference, the negative inverse Hessian of the logarithm of (3.3), evaluated at $\widehat{\boldsymbol{\theta}}$, can be used to approximate the large-sample covariance matrix of $\widehat{\boldsymbol{\theta}}$. The likelihood presented in (3.3) along with the proposed methods of estimation and inference are similar to the techniques outlined in Vansteelandt et al. (2000) and McMahan et al. (2013) with a few marked differences. In particular, through the dilution submodel h the proposed approach allows the sensitivity of the test to change from pool to pool, unlike Vansteelandt et al. (2000) which proceeds under the more traditional assumption that S_e is the same for all pools. By modeling the dilution effect, one can evaluate (3.3) based solely on the observed testing data, unlike the approach presented in McMahan et al. (2013) which requires additional a priori information on biomarkers to develop pool-specific testing accuracies.

Dorfman decoding

When classification is completed through Dorfman decoding, the observed data consist of $\mathcal{D}_D = \{\mathbf{Z}, \mathbf{Y}, \mathbf{X}\}$; i.e., the master pool responses, the individual retesting data observed from resolving positive pools, and the individual level covariates. To accommodate data of this structure into a regression analysis, it is assumed that the observed testing responses are independent, given the true statuses of the individuals. This assumption is common among the group testing literature when case identification is the goal; e.g., see Kim et al. (2007). Under this assumption, the observed

data likelihood can be expressed as

$$L(\boldsymbol{\theta}|\mathcal{D}_D) = \sum_{\widetilde{\mathbf{Y}}\in\widetilde{\boldsymbol{\mathcal{Y}}}} T_1(\boldsymbol{\beta}, \mathcal{D}_D, \widetilde{\mathbf{Y}}) T_2(\lambda, \mathcal{D}_D, \widetilde{\mathbf{Y}}) T_3(\mathcal{D}_D, \widetilde{\mathbf{Y}}), \tag{3.4}$$

where $\widetilde{\boldsymbol{\mathcal{Y}}}$ denotes the collection of all possible outcomes of $\widetilde{\mathbf{Y}}$, and

$$\begin{aligned}
T_1(\boldsymbol{\beta}, \mathcal{D}_D, \widetilde{\mathbf{Y}}) &= \prod_{j=1}^{J}\prod_{i=1}^{c_j} p_{ij}^{\widetilde{Y}_{ij}}(1-p_{ij})^{1-\widetilde{Y}_{ij}}, \\
T_2(\lambda, \mathcal{D}_D, \widetilde{\mathbf{Y}}) &= \prod_{j=1}^{J}\prod_{k=1}^{c_j} \left[h(k, c_j, \lambda)^{Z_j}\{1-h(k, c_j, \lambda)\}^{(1-Z_j)}\right]^{I_{jk}}, \\
T_3(\mathcal{D}_D, \widetilde{\mathbf{Y}}) &= \prod_{j=1}^{J}\{S_p^{(1-Z_j)}(1-S_p)^{Z_j}\}^{I_{j0}} \\
&\quad \times \left[\prod_{i=1}^{c_j}\{S_e^{\widetilde{Y}_{ij}}(1-S_p)^{1-\widetilde{Y}_{ij}}\}^{Y_{ij}}\{S_p^{1-\widetilde{Y}_{ij}}(1-S_e)^{\widetilde{Y}_{ij}}\}^{1-Y_{ij}}\right]^{Z_j},
\end{aligned}$$

where $I_{jk} = I(\sum_{i=1}^{c_j}\widetilde{Y}_{ij} = k)$, for $k = 1, 2, ..., c_j$. Due to the dimensionality of $\widetilde{\boldsymbol{\mathcal{Y}}}$, direct evaluation of (3.4) can be computationally burdensome.

To circumvent computational issues, we develop an expectation-maximization (EM) algorithm to find the MLE of $\boldsymbol{\theta}$ by viewing the individuals' latent statuses as "missing data." The complete data likelihood can be expressed as

$$L_c(\boldsymbol{\theta}|\mathcal{D}_D, \widetilde{\mathbf{Y}}) = T_1(\boldsymbol{\beta}, \mathcal{D}_D, \widetilde{\mathbf{Y}}) T_2(\lambda, \mathcal{D}_D, \widetilde{\mathbf{Y}}) T_3(\mathcal{D}_D, \widetilde{\mathbf{Y}}). \tag{3.5}$$

The E-step in the algorithm finds $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = E\{l_c(\boldsymbol{\theta}|\mathcal{D}_D, \widetilde{\mathbf{Y}})|\mathcal{D}_D, \boldsymbol{\theta}^{(d)}\}$, where the expectation is taken with respect to the individuals' latent statuses and $l_c(\boldsymbol{\theta}|\mathcal{D}_D, \widetilde{\mathbf{Y}})$ denotes the logarithm of (3.5). The form of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ can be expressed as

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)}) = \mathcal{T}_1(\boldsymbol{\beta}, \mathcal{D}_D, \boldsymbol{\theta}^{(d)}) + \mathcal{T}_2(\lambda, \mathcal{D}_D, \boldsymbol{\theta}^{(d)}) + \mathcal{T}_3(\mathcal{D}_D, \boldsymbol{\theta}^{(d)}),$$

where

$$\begin{aligned}
\mathcal{T}_1(\boldsymbol{\beta}, \mathcal{D}_D, \boldsymbol{\theta}^{(d)}) &= \sum_{j=1}^{J}\sum_{i=1}^{c_j} E(\widetilde{Y}_{ij}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})\log p_{ij} + \{1 - E(\widetilde{Y}_{ij}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})\}\log(1-p_{ij}) \\
\mathcal{T}_2(\lambda, \mathcal{D}_D, \boldsymbol{\theta}^{(d)}) &= \sum_{j=1}^{J}\sum_{k=1}^{c_j}\left[Z_j \log h(k, c_j, \lambda) + (1-Z_j)\log\{1-h(k, c_j, \lambda)\}\right] \\
&\quad \times E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)}),
\end{aligned}$$

and $\mathcal{T}_3(\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ is free of $\boldsymbol{\theta}$. We provide in Appendix B.1 closed-form expressions for both $E(\widetilde{Y}_{ij}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ and $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$. It is important to note that for large values of c_j (e.g., $c_j > 20$) evaluating these expectations directly can be computationally intractable. An alternate approach to approximate these expectations is provided in Appendix B.1.

The M-step of the algorithm finds $\boldsymbol{\theta}^{(d+1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$. Given the form of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(d)})$ above, the maximization step can be divided into two stages; i.e., one can find

$$\begin{aligned} \boldsymbol{\beta}^{(d+1)} &= \arg\max_{\boldsymbol{\beta}} \mathcal{T}_1(\boldsymbol{\beta}, \mathcal{D}_D, \boldsymbol{\theta}^{(d)}), \\ \lambda^{(d+1)} &= \arg\max_{\lambda} \mathcal{T}_2(\lambda, \mathcal{D}_D, \boldsymbol{\theta}^{(d)}), \end{aligned}$$

where $\boldsymbol{\theta}^{(d+1)} = (\boldsymbol{\beta}^{(d+1)\prime}, \lambda^{(d+1)})'$. The optimization step for $\boldsymbol{\beta}^{(d+1)}$ can be completed using any standard optimization routine appropriate for fitting binary regression models; e.g., `optim` in R. Similarly, because $\lambda \geq 0$ the optimization for λ is over a unidimensional constrained space and can be completed using standard numerical optimization routines; e.g., `optimize` in R.

In what follows, the EM algorithm is summarized. Initialize $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)\prime}, \lambda^{(0)})'$, set $d = 0$, and repeat the following steps until convergence.

1. Evaluate $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$, for $k = 1, 2, ..., c_j$ and $j = 1, 2, ..., J$.
2. Find $\lambda^{(d+1)} = \arg\max_{\lambda} \mathcal{T}_2(\lambda, \mathcal{D}_D, \boldsymbol{\theta}^{(d)})$.
3. Evaluate $E\{\widetilde{Y}_{ij}|\mathcal{D}_D, (\boldsymbol{\beta}^{(d)\prime}, \lambda^{(d+1)})'\}$, for $i = 1, 2, ..., c_j$ and $j = 1, 2, ..., J$.
4. Find $\boldsymbol{\beta}^{(d+1)} = \arg\max_{\boldsymbol{\beta}} \mathcal{T}_1(\boldsymbol{\beta}, \mathcal{D}_D, (\boldsymbol{\beta}^{(d)\prime}, \lambda^{(d+1)})')$.
5. Update $\boldsymbol{\theta}^{(d+1)} = (\boldsymbol{\beta}^{(d+1)\prime}, \lambda^{(d+1)})'$ and set $d = d + 1$.

At convergence of the algorithm, take $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(d)}$ to be the MLE of $\boldsymbol{\theta}$. By direct appeal to the missing data principle and the method outlined in Louis (1982), one can obtain

the observed information matrix as

$$\mathcal{I}(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \text{cov}\left\{ \frac{\partial l_c(\boldsymbol{\theta}|\mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \boldsymbol{\theta}} \middle| \mathcal{D}_D, \boldsymbol{\theta} \right\}.$$

Appendix B.1 provides closed-form expressions for both of the terms on the right hand side of $\mathcal{I}(\boldsymbol{\theta})$. Large-sample Wald inference can be conducted in the usual fashion using $\mathcal{I}(\widehat{\boldsymbol{\theta}})^{-1}$ as an estimate of the large-sample covariance matrix of $\widehat{\boldsymbol{\theta}}$.

3.3 Detecting the dilution effect

The dilution submodel h defined in Section 3.2 accounts for the presence or absence of dilution through the parameter λ; i.e., $\lambda > 0$ corresponds to the event that dilution is present, $\lambda = 0$ otherwise. Based on either $\mathcal{D}_M$ or $\mathcal{D}_D$, the modeling techniques outlined in Sections 3.2 and 3.2 can be used to obtain the MLE of λ, which is denoted as $\widehat{\lambda}$. Thus, through this parameter estimate, one gains subjective evidence of whether or not a dilution effect is present. In addition, a formal test can be constructed for the same purpose by considering the following set of hypotheses,

$$H_0 : \lambda = 0 \quad \text{and} \quad H_1 : \lambda > 0, \tag{3.6}$$

where rejecting H_0 suggests there is evidence of dilution. The practical relevance of this test is to assess whether the less complex model under H_0 (e.g., the model proposed by Vansteelandt et al., 2000) is appropriate for analyzing data observed from a group testing algorithm.

In order to test H_0 versus H_1, a likelihood ratio statistic is given by

$$T_{LR} = 2 \ln \left\{ \frac{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}|\mathcal{D})}{\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0} L(\boldsymbol{\theta}|\mathcal{D})} \right\}, \tag{3.7}$$

where $\boldsymbol{\Theta}_0$ denotes the constrained subset of the parameter space, $\boldsymbol{\Theta}$, specified under the null hypothesis and $\mathcal{D}$ denotes the observed data. Large values of T_{LR} provide evidence against the null hypothesis; i.e., of a significant dilution effect. Due to the

inherently complex nature of the likelihood functions presented in (3.3) and (3.4), providing exact finite sample critical values and or p-values for testing (3.6) through the test statistic in (3.7) appears to be intractable. Further, approximating the sampling distribution of T_{LR} through the use of standard large-sample theory is not appropriate under the null hypothesis because λ exists on the boundary of the parameter space under H_0. To circumvent these issues, the general result of Self and Liang (1987) can be used to establish that, asymptotically, $T_{LR} \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, where χ_0^2 is a random variable with a point mass distribution at 0 and χ_1^2 is a chi-squared random variable with 1 degree of freedom. Using this asymptotic result, one can calculate critical values and p-values in the usual fashion. Testing near or at the boundary of the parameter space is a commonly encountered problem in the random effects literature; e.g., see Self and Liang (1987) and Molenberghs and Verbeke (2007).

Evaluation of T_{LR} requires one to calculate the observed likelihood function $L(\boldsymbol{\theta}|\mathcal{D})$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. While $L(\boldsymbol{\theta}|\mathcal{D})$ is available in (3.3) for master pool testing, a direct evaluation of $L(\boldsymbol{\theta}|\mathcal{D})$ from (3.5) for Dorfman decoding can be practically infeasible because of the missing data involved. We show in Appendix B.3 how one can evaluate the likelihood function in an alternative approach.

3.4 Simulation evidence

We perform a simulation study to access the performance of our proposed estimation and testing approaches. We also compare our approach with the existing regression techniques that ignore dilution; for example, Vansteelandt et al. (2000) fits master pool testing data with constant S_e and S_p and Zhang et al. (2013) fits Dorfman decoding data with constant S_e and S_p.

Simulation description

We simulate $B = 500$ group testing data sets of sample size $N = 5000$ for both master pool testing and Dorfman decoding. Our simulation requires specification of the link function $g(\cdot)$ to generate individual true statuses and the submodel h to introduce dilution (i.e., false negative test results). First, we generate individual true statuses $\widetilde{Y}_{ij}$ using the logit link function $g(\cdot)$ given by

$$\text{logit}\{\text{pr}(\widetilde{Y}_{ij} = 1)\} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2},$$

where $x_{ij1} \sim \mathcal{N}(0, 0.75^2)$ and $x_{ij2} \sim$ Bernoulli(0.1), and $\boldsymbol{\beta} = (-3, 2, 1)'$. For this configuration, the individual-level disease probability is between $8\% - 10\%$ on average, which is consistent with the HBV data example we use in Section 3.5. Second, we assign individuals to pools using both random and homogeneous pooling strategies, as described in Vansteelandt et al. (2000). For random pooling, individuals are assigned to pools randomly. For homogeneous pooling, individuals are sorted first by x_{ij2} and then by x_{ij1} before they are assigned to pools. We choose the constant pool sizes $c_j = 5, 10$ and also a combination (UE) which consists of 334 pools of size 5 and 333 pools of size 10. Third, we determine true pooled statuses and then simulate group testing responses as $Z_j \sim \text{Bernoulli}\{h(k, c_j, \lambda)I_{jk} + S_p I_{j0}\}$, for $k = 1, 2, ..., c_j$, using the submodel

$$h(k, c_j, \lambda) = \frac{\exp\{\lambda\, \tau(k, c_j)\}}{S_e^{-1} + \exp\{\lambda\, \tau(k, c_j)\} - 1} \tag{3.8}$$

where $\tau_1(k, c_j) = (k - c_j)/c_j$ and $\lambda \geq 0$. For Dorfman decoding, individual retest results are simulated as $Y_{ij} \sim \text{Bernoulli}\{S_e\widetilde{Y}_{ij} + (1 - S_p)\widetilde{Y}_{ij}\}$. We use the individual-level testing accuracies $S_e = S_p = 0.99$. The submodel in (3.8) obeys all of the assumptions defined in Section 3.2 and adheres to the characteristics outlined in (3.1) and (3.2). We show in Appendix B.4 how we derive this submodel using a cumulative distribution function. We use $\lambda = 2.6, 3.8, 5.0$ to introduce mild, moderate, and severe misclassification. For these choices of λ, the dilution function h lies between

$85\% - 97\%$ on average. The values of h for these configurations are provided in Table 3.1. To avoid constrained optimization, we reparameterize as $\lambda = \exp(\phi)$ where $\phi \in (-\infty, \infty)$; that is, we estimate $(\beta_0, \beta_1, \beta_2, \phi)'$ instead of estimating $(\beta_0, \beta_1, \beta_2, \lambda)'$. In addition to increasing numerical stability, this transformation improves covariance matrix estimation for the MLE $\widehat{\boldsymbol{\theta}}$.

Table 3.1: Test sensitivity using the submodel h in (3.8) with different parameter configurations.

c_j	λ	$k=1$	2	3	4	5	6	7	8	9	10
	2.6	0.93	0.95	0.97	0.98	0.99	—	—	—	—	—
5	3.8	0.83	0.91	0.96	0.98	0.99	—	—	—	—	—
	5.0	0.64	0.83	0.93	0.97	0.99	—	—	—	—	—
	2.6	0.91	0.93	0.94	0.95	0.96	0.97	0.98	0.98	0.99	0.99
10	3.8	0.76	0.83	0.87	0.91	0.94	0.96	0.97	0.98	0.99	0.99
	5.0	0.52	0.64	0.75	0.83	0.89	0.93	0.96	0.97	0.98	0.99

Simulation results

We view the dilution parameter λ as a nuisance parameter and present estimation results only for the regression parameter $\boldsymbol{\beta}$. In Table 3.2, we present estimation results calculated from $B = 500$ simulated data sets for random pooling. We report averaged MLEs, averaged standard error estimates, and estimated coverage probabilities for large-sample 95% Wald confidence intervals. Overall, our dilution method provides estimates that are on target, on average, whereas techniques that ignores false negative test results can provide estimates that are severely biased. Furthermore, as the pool size becomes larger or the level of misclassification increases, the amount of bias becomes more pronounced. While our approach fixes estimation bias, it can yield less precise estimates of $\boldsymbol{\beta}$ because it (i) accounts for testing errors and (ii) additionally estimates λ. This phenomenon is mainly observed for master pool testing, although the loss of efficiency does not appear to be large. The dilution method works ex-

ceedingly well for Dorfman decoding; the MLEs are nearly as precise as the MLEs using the constant method in Zhang et al. (2013) which estimates only $\boldsymbol{\beta}$. This happens because the accuracy for individual retest results are assumed to be known ($S_e = S_p = 0.99$).

We find from Table 3.2 that failure to account for dilution causes existing techniques to drastically underestimate the nominal coverage rate for 95% Wald confidence intervals. This mainly results from the high amount of bias in the estimates of $\boldsymbol{\beta}$. As λ increases, underestimation becomes more evident. In contrast, estimated coverage rates using our approach are mostly on target for all parameters in $\boldsymbol{\beta}$. We show simulation evidence for homogeneous pooling in Appendix B.5. Estimation accuracy and precision for homogeneous pooling are similar to that for random pooling. This finding is somewhat counterintuitive given the evidence in Vansteelandt et al. (2000) who showed that homogeneous pooling produces more efficient estimates of $\boldsymbol{\beta}$.

For the estimation results discussed above, we correctly specified the true submodel in (3.8). One may wonder how robust these results are to dilution submodel misspecification. To investigate this issue, we performed a simulation in Appendix B.5 wherein we assume the submodel in (3.8) for the group testing data simulated using the three submodels in Appendix B.4. We found that estimation results are largely unaffected by submodel misspecification.

Finally, satisfactory performance of our approach requires large data sets. This is especially crucial for master pool testing. With smaller sample sizes such as in the HBV data application in Section 3.5, we do not advise to use the dilution model for master pool testing. However, if individual retest results are available from a screening algorithm, such as Dorfman decoding, halving algorithm, array testing, etc., the dilution approach should work nearly as well as individual testing; we do not show the comparison with individual testing for brevity. We sometimes experienced computational difficulties (especially for master pool testing) when estimating the

large-sample covariance matrix for $\widehat{\boldsymbol{\theta}}$. The estimated variance for $\widehat{\lambda}$ occasionally becomes negative. This problem mainly arises when $\widehat{\lambda}$ is close to zero (boundary of the parameter space). However, this problem can be resolved when the observed data consists of a sufficiently large number of test results.

Table 3.2: Simulation results for master pool testing (MPT) and Dorfman decoding (DD) with $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \lambda)' = (-3, 2, 1, \lambda)'$. "Mean" is the averaged maximum likelihood estimate and SE is the averaged standard error estimate calculated from 500 simulated data sets. Cov is the estimated coverage rate of nominal 95% Wald confidence intervals. The margin of error for the estimated coverage rate, assuming a 99% confidence level, is 0.03. Constant pool sizes c are used. Random pooling has been used for this simulation.

			Constant S_e/S_p			Dilution		
c			$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
				$\lambda = 2.6$				
5	MPT	Mean (SE)	−3.05 (0.13)	1.92 (0.15)	0.93 (0.26)	−2.98 (0.17)	2.07 (0.24)	1.01 (0.30)
		Cov	0.96	0.91	0.95	0.97	0.95	0.97
	DD	Mean (SE)	−3.06 (0.09)	1.97 (0.09)	0.97 (0.15)	−3.00 (0.10)	2.01 (0.10)	1.00 (0.16)
		Cov	0.91	0.92	0.91	0.93	0.94	0.92
10	MPT	Mean (SE)	−3.05 (0.18)	1.81 (0.23)	0.85 (0.40)	−3.07 (0.26)	2.21 (0.49)	1.03 (0.55)
		Cov	0.95	0.87	0.97	0.97	0.97	0.97
	DD	Mean (SE)	−3.07 (0.09)	1.96 (0.10)	0.97 (0.15)	−3.00 (0.10)	2.01 (0.10)	0.99 (0.16)
		Cov	0.89	0.93	0.96	0.96	0.96	0.96
				$\lambda = 3.8$				
5	MPT	Mean (SE)	−3.15 (0.13)	1.82 (0.15)	0.90 (0.26)	−3.01 (0.21)	2.08 (0.27)	1.05 (0.32)
		Cov	0.81	0.77	0.94	0.92	0.94	0.95
	DD	Mean (SE)	−3.14 (0.09)	1.92 (0.09)	0.95 (0.15)	−3.01 (0.10)	2.00 (0.10)	1.00 (0.16)
		Cov	0.64	0.84	0.95	0.94	0.95	0.96
10	MPT	Mean (SE)	−3.18 (0.17)	1.59 (0.22)	0.79 (0.41)	−3.10 (0.36)	2.29 (0.63)	1.15 (0.69)
		Cov	0.85	0.53	0.96	0.91	0.91	0.96
	DD	Mean (SE)	−3.21 (0.09)	1.90 (0.10)	0.93 (0.16)	−3.01 (0.10)	2.00 (0.11)	0.99 (0.17)
		Cov	0.41	0.83	0.93	0.94	0.94	0.95
				$\lambda = 5.0$				
5	MPT	Mean (SE)	−3.35 (0.13)	1.68 (0.16)	0.83 (0.28)	−3.03 (0.31)	2.04 (0.30)	1.02 (0.37)
		Cov	0.24	0.46	0.94	0.89	0.91	0.96
	DD	Mean (SE)	−3.34 (0.10)	1.84 (0.10)	0.89 (0.17)	−3.01 (0.11)	2.01 (0.12)	0.99 (0.18)
		Cov	0.04	0.61	0.91	0.95	0.94	0.96
10	MPT	Mean (SE)	−3.55 (0.18)	1.45 (0.23)	0.67 (0.47)	−3.09 (0.54)	2.19 (0.64)	1.10 (0.77)
		Cov	0.09	0.30	0.98	0.82	0.90	0.96
	DD	Mean (SE)	−3.50 (0.11)	1.82 (0.11)	0.89 (0.18)	−3.01 (0.12)	2.01 (0.12)	1.00 (0.20)
		Cov	0.00	0.59	0.92	0.96	0.96	0.94

Table 3.3: Estimated size and power of the $\alpha = 0.05$ likelihood ratio test calculated from 500 simulated data sets with $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \lambda)' = (-3, 2, 1, \lambda)'$. The margin of error for the estimated size when $\lambda = 0$, assuming a 99% confidence level, is 0.03. Constant pool sizes c and unequal (UE) pool sizes are used.

	Master pool testing					Dorfman decoding				
c	$\lambda = 0$	1	2	3	4	0	1	2	3	4
Random pooling										
5	0.05	0.08	0.06	0.13	0.22	0.05	0.07	0.16	0.51	0.93
10	0.05	0.09	0.10	0.20	0.27	0.05	0.10	0.32	0.91	1.00
UE	0.04	0.05	0.17	0.30	0.59	0.04	0.08	0.27	0.83	1.00
Homogeneous pooling										
5	0.04	0.06	0.08	0.14	0.16	0.06	0.06	0.17	0.54	0.92
10	0.06	0.09	0.20	0.31	0.38	0.03	0.10	0.34	0.78	1.00
UE	0.07	0.10	0.19	0.29	0.38	0.04	0.13	0.32	0.84	1.00

Power of the hypothesis test for dilution effects

To exhibit power properties of the likelihood ratio test in Section 3.3, we continue with the data simulation configurations described in Section 3.4 with one exception; we now consider $\lambda = 0, 1, ..., 4$. The estimated size and power of the $\alpha = 0.05$ likelihood ratio test is presented in Table 3.3. We estimate the power as the proportion of times $H_0 : \lambda = 0$ is rejected out of 500 data sets. As one might expect, the power of the test increases as λ increases. For all configurations, the size of the test is estimated well and is within the margin of error for $\alpha = 0.05$. The test for Dorfman decoding is much more powerful because it involves more test results. The overall performance of the test using variable pool sizes (UE) is better. This may be the result of the induced between-pool variability (in regard to testing sensitivities) from unequal group composition. Like the estimation results in Section 3.4, the power calculation results here are unaffected by the pooling strategies (random or homogeneous). We also explored the power properties when the submodel in (3.8) is misspecified (see Appendix B.5). With the misspecified submodels, the likelihood ratio test performs as well as with the correctly specified submodel; the size is estimated well and the power has a

monotonically increasing trend as the amount of dilution increases.

3.5 Data application

We illustrate our regression techniques using HBV data from a national study conducted in Ireland and reported in Allwright et al. (2000). The data consist of HBV test results, OD readings from a Murex ICE enzyme immunoassay, and covariate information, such as age, drug use, and sexual practices, for $N = 1193$ Irish prisoners. HBV tests were performed individually using oral fluid specimens and the covariates were collected through a voluntary survey performed anonymously. The goal of this study was to estimate the overall prevalence of positivity and to identify risk factors for infection.

To incorporate false negative results in group testing, we adopt the strategy demonstrated in McMahan et al. (2013). We use age as a covariate and HBV disease statuses and OD readings to determine testing responses. In our analysis, we use 1137 individuals (99 positive cases and 1038 negative cases) for whom we have complete information. The variable age ranges from 16 to 67 years, 91% of the individuals are aged between 16 and 40, and multiple individuals share the same age. We consider both random and homogeneous pooling strategies. For homogeneous pooling, individuals are sorted by age before assigning to pools.

We assume the OD readings are measured without error because we do not have any additional information about these observations. First, we assign individuals to pools and then calculate the pooled OD reading as $\mathrm{OD}_j = c_j^{-1} \sum_{i=1}^{c_j} \mathrm{OD}_{ij}$. This proceeds under the assumption adopted by McMahan et al. (2013) that the OD reading for a pool is the average of the individuals' OD readings in the pool. To determine diagnosed statuses, we find the cutoff t^* (which is not available to us) that

minimizes the discrepancies between the individuals' diagnosed statuses so that

$$t^* = \arg\min_t \left\{ \sum_{i=1}^{N^+} I(\mathrm{OD}_i^+ < t) + \sum_{i=1}^{N^-} I(\mathrm{OD}_i^- > t) \right\},$$

where $N^+ = 99$, $N^- = 1038$, and OD_i^+ and OD_i^- are the OD readings observed from HBV-positive and HBV-negative individuals. Next, we determine pool testing responses by $Z_j = I(\mathrm{OD}_j > t^*)$ and individual retest results by $Y_{ij} = I(\mathrm{OD}_{ij} > t^*)$. Recall that, for Dorfman decoding, the retest results $\{Y_{1j}, Y_{2j}, ..., Y_{c_j j}\}$ are observed only when the jth pool is diagnosed as positive. For simplicity, we use constant pool sizes $c_j = 1, 3, 4, 5, 6, 8, 10$, where $c_j = 1$ corresponds to individual testing. When $N/c_j > 0$, we form one reminder pool of smaller size. We repeat this entire procedure $B = 500$ times to obtain 500 group testing data sets.

For individual-level disease probabilities, we consider the following logistic models

$$\mathrm{logit}\{\mathrm{pr}(\widetilde{Y}_{ij} = 1)\} = \beta_0 + \beta_1 x_{ij} \tag{3.9}$$

$$\mathrm{logit}\{\mathrm{pr}(\widetilde{Y}_{ij} = 1)\} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2, \tag{3.10}$$

where the covariate $x = (\mathrm{age} - \overline{\mathrm{age}})/SD$, $\overline{\mathrm{age}}$ is the mean age and SD is the standard deviation for age. We found that standardization improves numerical stability for the polynomial model in (3.10). Because we treat OD readings as being measured without error, the individual testing accuracies are calculated as

$$S_e = \frac{1}{N^+} \sum_{i=1}^{N^+} I(\mathrm{OD}_i^+ > t^*) \quad \text{and} \quad S_p = \frac{1}{N^-} \sum_{i=1}^{N^-} I(\mathrm{OD}_i^- < t^*).$$

For this small sample study, we fit the dilution model only for Dorfman decoding, and we continue to use the submodel h given by (3.8). To make our comparisons, we also fit an individual testing model and the model in Zhang et al. (2013) which assumes S_e and S_p to be constant.

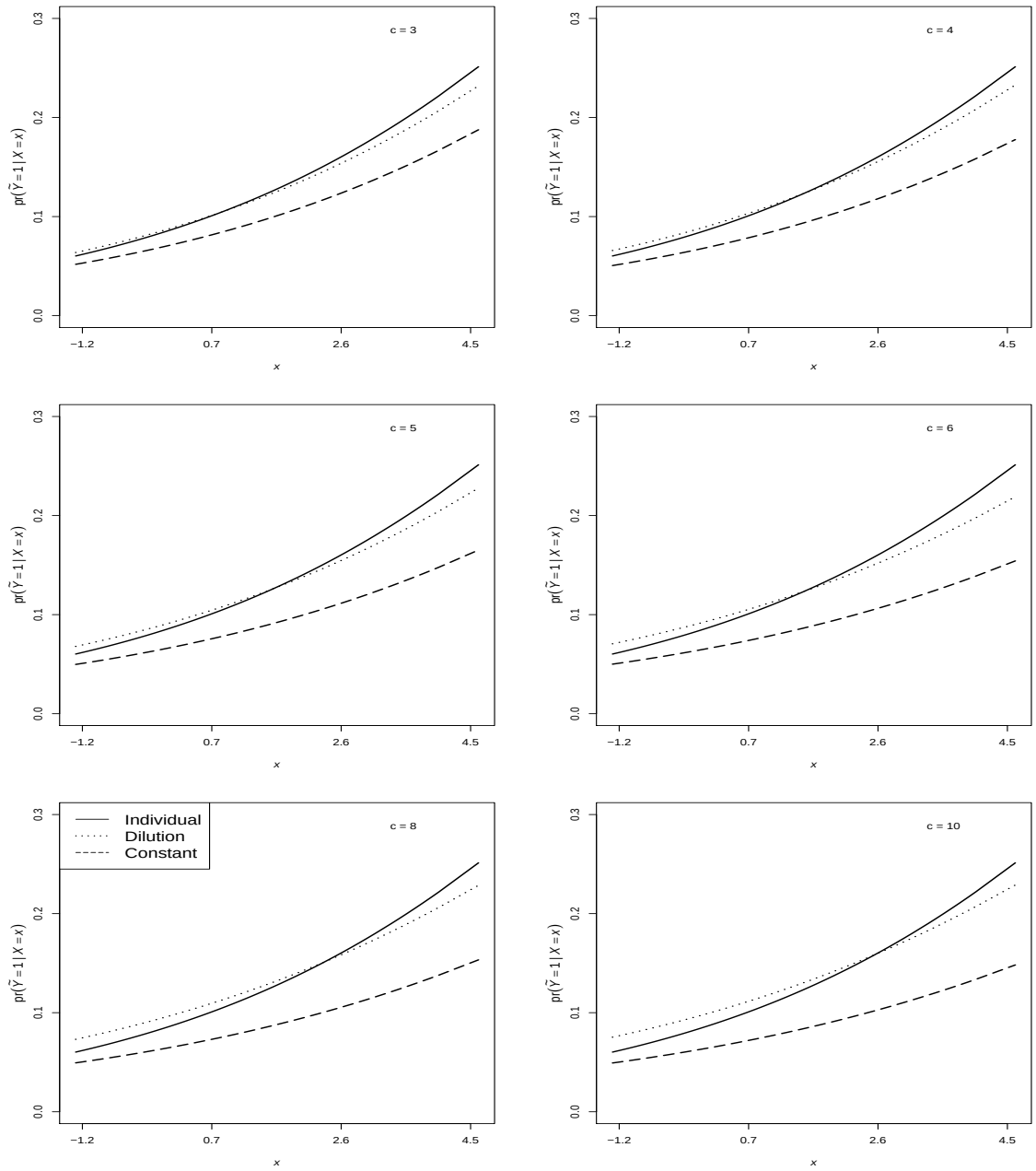

Figure 3.1: Irish HBV data analysis with Dorfman decoding and random pooling. The first-order logistic model in (3.9) is assumed. Estimated regression functions, averaged over $B = 500$ implementations, are presented. The estimated regression function for individual testing is also shown for comparison.

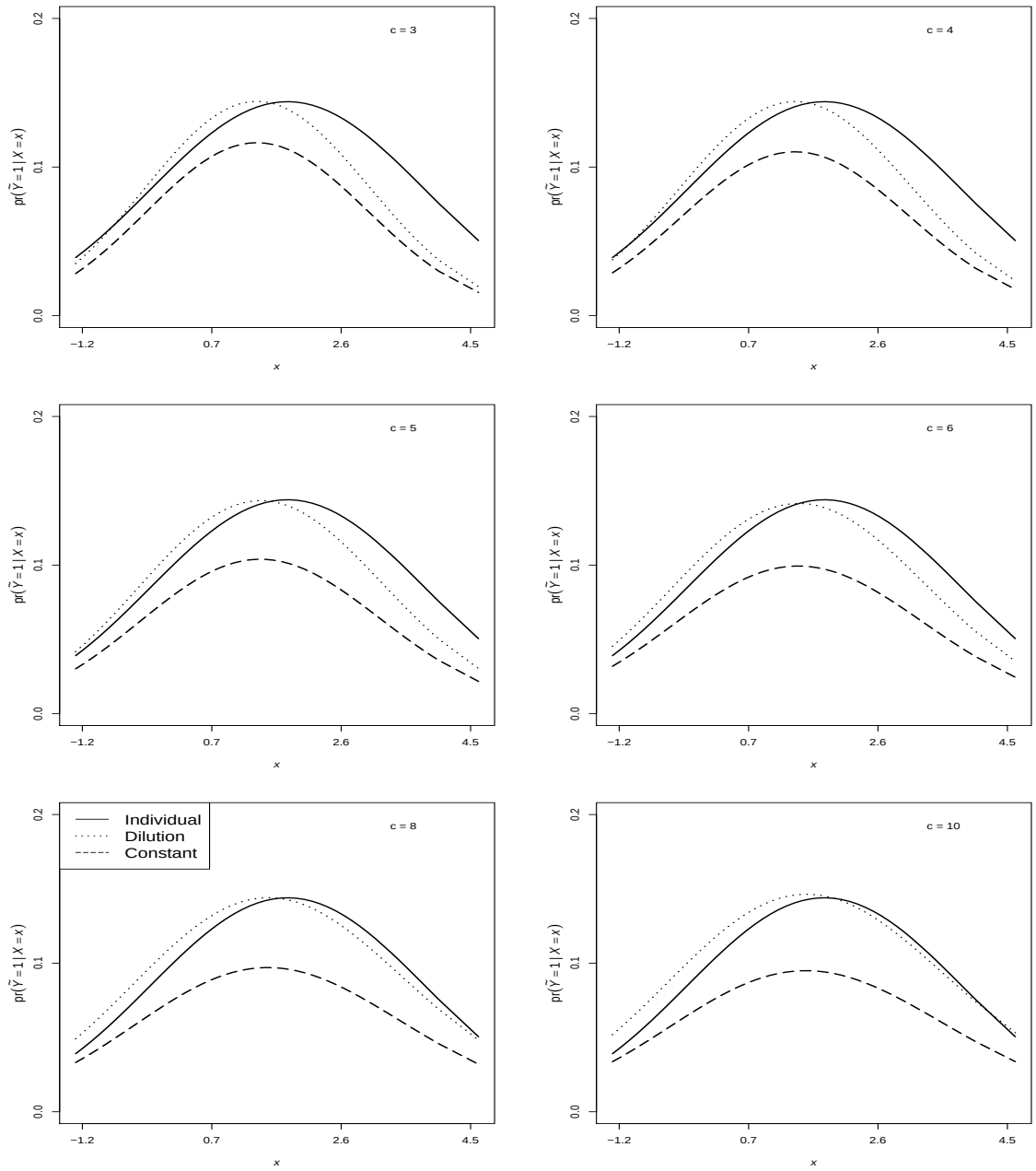

Figure 3.2: Irish HBV data analysis with Dorfman decoding and random pooling. The polynomial logistic model in (3.10) is assumed. Estimated regression functions, averaged over $B = 500$ implementations, are presented. The estimated regression function for individual testing is also shown for comparison.

Table 3.4: Irish HBV data analysis with Dorfman decoding. The first-order logistic model in (3.9) is assumed. MLE (estimated standard error) for $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ averaged over $B = 500$ implementations. "Reject" is the proportion that the likelihood ratio test in Section 3.3 detects dilution using the level of significance α. Individual testing ($c = 1$) estimates are also reported for comparison.

	Constant S_e/S_p		Dilution		Reject	
c	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\alpha = 0.05$	0.10
1	−2.38 (0.11)	0.28 (0.09)	—	—	—	—
			Random pooling			
3	−2.59 (0.12)	0.24 (0.10)	−2.37 (0.20)	0.25 (0.11)	0.30	0.42
4	−2.63 (0.12)	0.24 (0.10)	−2.34 (0.19)	0.25 (0.11)	0.49	0.66
5	−2.66 (0.12)	0.22 (0.11)	−2.32 (0.18)	0.24 (0.11)	0.69	0.79
6	−2.67 (0.12)	0.21 (0.11)	−2.30 (0.18)	0.22 (0.11)	0.81	0.87
8	−2.69 (0.12)	0.21 (0.11)	−2.26 (0.17)	0.22 (0.11)	0.93	0.97
10	−2.70 (0.13)	0.20 (0.11)	−2.23 (0.16)	0.22 (0.12)	0.98	0.99
			Homogeneous pooling			
3	−2.59 (0.12)	0.25 (0.10)	−2.35 (0.20)	0.24 (0.11)	0.35	0.50
4	−2.62 (0.12)	0.25 (0.11)	−2.32 (0.19)	0.25 (0.11)	0.57	0.72
5	−2.65 (0.12)	0.23 (0.11)	−2.29 (0.18)	0.22 (0.11)	0.76	0.88
6	−2.66 (0.12)	0.22 (0.11)	−2.25 (0.17)	0.21 (0.12)	0.90	0.97
8	−2.67 (0.12)	0.22 (0.11)	−2.20 (0.16)	0.22 (0.12)	0.99	0.99
10	−2.67 (0.13)	0.25 (0.11)	−2.18 (0.16)	0.23 (0.12)	1.00	1.00

Figure 3.1 presents averaged estimates of the first-order regression function in (3.9). Table 3.4 shows averaged MLEs, averaged standard error estimates, and the proportion of times the likelihood ratio test in Section 3 detects dilution, for the regression function in (3.9). These estimation results are calculated from the 500 data sets using random pooling. One observes that the performance of the dilution regression method is close to that for individual testing, despite the fact that the former is estimating the additional parameter λ with significantly fewer number of test results. On the other hand, the constant S_e/S_p method in Zhang et al. (2013) provides inaccurate estimates and performs worse as more dilution is incorporated by increasing pool sizes. These results reinforce the simulation evidence in Section 3.4.

One can make several remarks from the hypothesis test results in Table 3.4. The rejection rate increases with c_j as one would expect. Homogeneous pooling appears to help detect the dilution, although this outcome is not observed from the simulation in Section 3.4. Also, changing the level of significance from $\alpha = 0.05$ to $\alpha = 0.10$ shows a noticeable increase in the rejection rates, as expected. Finally, Figure 3.2 shows estimated regression functions, averaged over 500 data sets, for the polynomial model in (3.10). These results are fairly consistent with the results discussed above. However, the regression functions are slightly underestimated in the extremes when c_j is smaller, perhaps because only a few individuals are older than 40. More results from this data analysis are presented in Appendix B.6.

3.6 Discussion

We have generalized group testing regression models to account for misclassification due to dilution effects. Our approach corrects for estimation bias by exploiting a parametric function specified for pooled-level sensitivity and provides reliable inference. While McMahan et al. (2013) studied the same problem (for master pool testing), one of the notable advantages of our framework is that a formal hypothesis test can be performed to detect dilution. This enables one to decide whether to fit our complex dilution model which accounts for false negative test results.

We have assumed that pool testing sensitivities can be modeled by a parametric function h whose form is known. Such an assumption is not overly restrictive. One can select h using pilot study data or an assay's validation data. In the absence of such data, one can construct the function using any cumulative distribution function as we have shown in Appendix B.4. While any submodel which possesses the characteristics of h as defined in Section 3.2 can be used, careful choice of h will ensure better estimation precision. Our dilution approach requires a sufficient amount of data. This is not unrealistic because group testing is often used for large screening applications

(e.g., chlamydia and gonorrhea screening by CDC and HIV, HBV, and HCV screening by American Red Cross). For example, the state of Nebraska alone tests 20-30 thousand individuals every year for chlamydia and gonorrhea.

In this work, we have assumed that individual testing sensitivity and specificity (S_e and S_p) are known. Future work might treat these quantities as unknown and make efforts to estimate them from the observed data. This would require even larger data sets to identify the submodel model h, in which case a Bayesian approach might be more suitable. One can construct informative prior distributions using the plethora of prior information available in most disease screening applications. Finally, we have generalized the regression models in Vansteelandt et al. (2000) and Zhang et al. (2013) for master pool testing and Dorfman decoding. It would be straightforward to generalize our methods to handle other group testing strategies, such as array based testing and the halving algorithm in Zhang et al. (2013). One could also generalize our approach to be used with other types of models, such as the random effects model described in Chen, Tebbs, and Bilder (2009).

Chapter 4

Group testing regression with measurement error in covariates

4.1 Introduction

The seminal work in group testing was motivated by a large-scale infection screening application (Dorfman, 1943). During World War II, the United States Public Health Service and associated organizations used to test inductees to weed out the ones infected by syphilis. To accomplish such a massive testing task with affordable costs and efforts, Dorfman proposed to test the inductees through pooled blood specimens as an alternative to the conventional one-by-one testing. Since then, group testing, also called pooled testing, has been effectively used in applications involving sexually transmitted diseases, such as HIV (Pilcher et al., 2005), chlamydia and gonorrhea (Lewis et al., 2012), and HBV and HCV (Hourfar et al., 2008; Stramer et al., 2013). Group testing can be successfully implemented in any applications when two primary requirements are met: (i) trait of interest (e.g., defective/non-defective) is rare, (ii) pools can be formed by compositing a set of individual specimens. For group testing applications in other areas see, animal testing (Dhand et al., 2010), genetics (Chi et al., 2009), pollution detection (Wahed et al., 2006), food safety (Fahey et al., 2006), and drug discovery (Remlinger et al., 2006).

Statistical research in group testing is mainly found in a homogeneous population setting, where the goal is to estimate a binomial proportion (Mendoza-Blanco et al., 1996; Johnson and Pearson, 1999; Hanson et al., 2006). In recent years, group testing

research has shifted towards heterogeneous population settings where individual-level covariates can be useful in both classification (case identification) and estimation problems. To improve efficiency in classification, Bilder et al. (2010) and McMahan et al. (2012a,b) used covariates in the design stage of group testing. In estimation problems, covariate-adjusted inference has been proposed in all research directions: parametric (Vansteelandt et al., 2000; Bilder and Tebbs, 2009), semiparametric (Wang et al., 2014b), and nonparametric (Delaigle and Meister, 2011; Delaigle and Zhou, 2015). Furthermore, group testing regression models have been studied for more sophisticated data structures. Chen et al. (2009) proposed random effects models to account for regional variability, and McMahan et al. (2013), Wang et al. (2015) and Delaigle and Hall (2015) proposed regression models in the presence of dilution effects.

Accuracy in group testing regression inference can be compromised by two basic sources of error contamination: (i) errors in testing responses, (ii) errors in covariate measurements. While the former has been studied widely, the later is still very new and a fulfilling research direction. Measurement errors in covariates often arise in epidemiological applications. For example, subjects infected by one sexually transmitted disease, such as chlamydia or gonorrhea, are susceptible to another disease such as HIV. Therefore, to calculate a subject's disease probability for HIV, it is crucial to use the subject's infection status for chlamydia or gonorrhea as a covariate. Because these types of covariates are often measured with errors, it is necessary to acknowledge such errors to accurately calculate covariate-specific disease probabilities. Huang and Tebbs (2009) first studied measurement error models for group testing data. These authors developed a diagnostic tool that can identify misspecification in structural measurement error models. Huang (2009) proposed an improved version of the diagnostic method in Huang and Tebbs (2009). These articles did not focus on estimation and are limited to structural measurement error models, which require a

known probability distribution for true (latent) covariates. Misspecification of such latent model can adversely affect inference (Carroll et al., 2006; Huang and Tebbs, 2009). Furthermore, these articles assume the availability of a perfect assay. This requirement can be too prohibitive for most applications in group testing. Delaigle and Meister (2011) briefly talked about a nonparametric approach to model group testing data in the presence of measurement errors.

We take a Bayesian approach to generalize the existing group testing models in the presence of measurement errors. Our method offers a number of appealing features. First, we treat the latent covariates as "missing data" on which a flexible prior distribution is specified. Moreover, to emulate a study where no such prior information is available, we estimate the latent model nonparametrically. Second, our approach provides flexibility to the error structure which relates observed covariates to latent covariates, unlike most existing approaches which assume the measurement error variance fixed and known (Huang and Tebbs, 2009; Huang, 2009). Third, we allow for imperfect diagnoses and incorporate information about the assay's uncertainty from assay product literature. Finally, one can construct sound informative priors using the abundance of historical data available in epidemiological applications where group testing is often used. In the next section, we formulate our model and discuss how the inference can be performed via Markov Chain Monte Carolo (MCMC) techniques.

4.2 Model formulation

We consider a group testing application of N individuals who are randomly assigned to one of the J non-overlapping pools. Let $\widetilde{Y}_{ij}$ denote true disease statuses, where $\widetilde{Y}_{ij} = 1$ if the ith individual in the jth is truly positive and $\widetilde{Y}_{ij} = 0$ if otherwise, for $i = 1, 2, ..., c_j$ and $j = 1, 2, ..., J$. Let c_j denote the pool size and $N = \sum_{j=1}^{J} c_j$, where $c_j = 1$ refers to individual testing. We assume that $\widetilde{Y}_{ij} \sim$ Bernoulli(p_{ij}); i.e., $\widetilde{Y}_{ij}$'s are independent Bernoulli random variables with mean p_{ij}. Denote by X_{ij} the true (scalar) covariates of the ith individual in the jth pool. In this study, X_{ij}'s are unobserved and are regarded as missing data. For ease of exposition, we develop the model with the univariate covariates X_{ij}. Extension of our model to multivariate settings will be straightforward. Let individuals' disease statuses relate to their true covariates as

$$p_{ij} = \text{pr}(\widetilde{Y}_{ij} = 1 | X_{ij}, \boldsymbol{\beta}) = g^{-1}(\beta_0 + \beta_1 X_{ij}), \tag{4.1}$$

where $g(\cdot)$ is a monotone and differentiable link function and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is the vector of regression parameters. Inference about $\boldsymbol{\beta}$ is of central interest.

Suppose W_{ij}'s are observed covariates that are prone to measurement error; i.e., we observe the error contaminated covariates W_{ij} in place of X_{ij}. We assume an additive error model structure which links W_{ij} to X_{ij} as

$$W_{ij} = X_{ij} + U_{ij}, \tag{4.2}$$

where U_{ij}'s are independent errors and $U_{ij} | \sigma_U^2 \sim \mathcal{N}(0, \sigma_U^2)$. Thus, $W_{ij} | \{X_{ij}, \sigma_U^2\} \sim \mathcal{N}(X_{ij}, \sigma_U^2)$. Note, we treat σ_U^2 as an unknown random variable. In the presence of replicate measurements on X_{ij}, one can estimate σ_U^2 and can treat σ_U^2 as a known constant. However, this assumption can be questionable when sufficiently large number of replicates are unavailable.

Let $\widetilde{Z}_j$ denote true pool statuses, where $\widetilde{Z}_j = 1$ if the jth pool contains at least one positive individual and $\widetilde{Z}_j = 0$ if otherwise. Denote the pool testing responses by Z_j, where $Z_j = 1$ if the jth pool is diagnosed as positive and $Z_j = 0$ if otherwise. In the presence of testing errors, $\widetilde{Z}_j$'s can not be unobserved, and one observes Z_j instead of $\widetilde{Z}_j$. Let the misclassification in testing responses be governed by assay sensitivity and specificity defined as $S_e = \text{pr}(Z_j = 1|\widetilde{Z}_j = 1)$ and $S_p = \text{pr}(Z_j = 0|\widetilde{Z}_j = 0)$. We assume that S_e and S_p do not depend on the pool size c_j. This assumption is common in group testing (Kim et al., 2007) and is reasonable when an assay's threshold is carefully chosen to accommodate both pooled and individual specimens. Let $\boldsymbol{\delta} = (S_e, S_p)'$. The probability that a pool tests positively is

$$\text{pr}(Z_j = 1|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta}) = S_e - (S_e + S_p - 1) \prod_{i=1}^{c_j} \left\{1 - g^{-1}\left(\beta_0 + \beta_1 X_{ij}\right)\right\},$$

where $\mathbf{X}_j = (X_{1j}, X_{2j}, ..., X_{c_jj})'$. Define $\mathbf{W}_j = (W_{1j}, W_{2j}, ..., W_{c_jj})'$. We assume that Z_j and $\mathbf{W}_j$, for $j = 1, 2, ..., c_j$, are independent conditional on the true covariates $\mathbf{X}_j$; i.e., $\{Z_j|\mathbf{X}_j\} \perp \{\mathbf{W}_j|\mathbf{X}_j\}$. This assumption is analogous to the "nondifferentiality assumption" commonly arises in individual testing regression (Carroll et al., 2006; Huang and Tebbs, 2009). The joint distribution of $(Z_j, \mathbf{W}_j)$ conditional on $\mathbf{X}_j$ is given by

$$f(Z_j, \mathbf{W}_j|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2) = f(Z_j|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta}) f(\mathbf{W}_j|\mathbf{X}_j, \sigma_U^2),$$

where

$$\begin{aligned} f(\mathbf{W}_j|\mathbf{X}_j, \sigma_U^2) &= \prod_{i=1}^{c_j} \sigma_U^{-1} \phi\{\sigma_U^{-1}(W_{ij} - X_{ij})\} \\ f(Z_j|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta}) &= \text{pr}(Z_j = 1|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta})^{Z_j} \{1 - \text{pr}(Z_j = 1|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta})\}^{1-Z_j}, \end{aligned}$$

and $\phi(\cdot)$ denotes the probability density function of a standard normal random variable. Let $\mathbf{Z} = (Z_1, Z_2, ..., Z_J)'$, $\mathbf{W} = (\mathbf{W}_1{}', \mathbf{W}_2{}', ..., \mathbf{W}_J{}')'$ and $\mathbf{X} = (\mathbf{X}_1{}', \mathbf{X}_2{}', ..., \mathbf{X}_J{}')'$. Thus, the joint distribution of the observed data $(\mathbf{Z}, \mathbf{W})$ given $(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2)$ is

$$f(\mathbf{Z}, \mathbf{W}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2) = \prod_{j=1}^{J} f(Z_j, \mathbf{W}_j|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2). \tag{4.3}$$

Having derived the observed data likelihood function in (4.3), we are left to specify a prior distribution $f(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2)$ for the unknown quantity $(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2, \mathbf{X})$, which is treated as random in a Bayesian framework. It follows that the posterior distribution is given by

$$\pi(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2 | \mathbf{Z}, \mathbf{W}) \propto f(\mathbf{Z}, \mathbf{W} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2) f(\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2). \tag{4.4}$$

We specify commonly used prior distributions. For the regression parameter $\boldsymbol{\beta}$, we elicit a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; i.e., $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We specify independent beta prior distributions for the test accuracy parameters in $\boldsymbol{\delta}$, i.e., $S_e \sim \text{beta}(a_{S_e}, b_{S_e})$ and $S_p \sim \text{beta}(a_{S_p}, b_{S_p})$. This independent beta-prior approach is analogous to the approach in Johnson and Pearson (1999) and Hanson et al. (2006). We choose inverse-gamma prior for the error variance, i.e., $\sigma_U^2 \sim \text{IG}(a_{\sigma_u^2}, b_{\sigma_u^2})$. Note, the aforementioned hyperparameters are assumed known constant. For the latent variables in $\mathbf{X}$, we use a hierarchical model structure; i.e., $X_{ij} | \{\mu_X, \sigma_X^2\} \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $\mu_X \sim \mathcal{N}(\mu_{\mu_x}, \sigma_{\mu_x}^2)$, and $\sigma_X^2 \sim \text{IG}(a_{\sigma_x^2}, a_{\sigma_x^2})$. Note, specification of the prior for X_{ij} should be close to the true distribution of X_{ij}. Assuming mutual independence among the parameters in $(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma_U^2, \mathbf{X})$, one can write out the posterior distribution in (4.4) and derive full conditional distributions. Inference about these parameters can be made conventionally using MCMC samples from those conditional distributions. Commonly used sampling techniques are Gibbs sampler (Geman and Geman, 1984) and Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) corresponding to the situations whether closed-form expressions for these condition distributions are available or not.

In the event that subjective knowledge about the true covariates X_{ij} is hard to obtain or is completely unavailable, we suggest to model X_{ij} nonparametrically, instead of specifying the hierarchical prior described above. It is straightforward to generalize our work to the case when replicate measurements are available on X_{ij}. In what follows, we provide a brief description of the generalization. Let

$\mathbf{W}^*_{ij} = (W_{ij1}, W_{ij2}, ..., W_{ijK})'$ denote the vector of K measurements for the ith subject in the jth pool. Because $\mathbf{W}^*_{ij}$ is obtained from the same subject, we treat the observations in $\mathbf{W}^*_{ij}$ correlated. Continuing with the assumptions about the additive error model in (4.2), we have $\mathbf{W}^*_{ij}|\{X_{ij}, \mathbf{\Sigma}_{W^*}\} \sim \mathcal{N}(X_{ij}\mathbf{1}_K, \mathbf{\Sigma}_{W^*})$, a multivariate normal distribution. To obtain the Bayesian model in presence of replicate measurements, one can replace the univariate model for $W_{ij}|\{X_{ij}, \sigma^2_U\}$ by the multivariate model for $\mathbf{W}^*_{ij}|\{X_{ij}, \mathbf{\Sigma}_{W^*}\}$ in (4.4).

In this chapter, we have outlined a Bayesian regression model in the presence of measurement errors in covariates (and also in testing responses). Our model can be viewed as a generalization of the existing regression models in group testing. Note, in the absence of replicate measurements on X_{ij}, the observed data do not provide much information about the error variance σ^2_U. Hence, one will need to specify a heavily informative prior for σ^2_U. Such a requirement can be weakened when replicate measurements are available. I have not explored this project completely. To access performance of the proposed model, I plan to do an extensive simulation study in future. I also plan to apply this model to the HBV data collected from the Irish prison population (Allwright et al., 2000). Description about the Irish HBV data is provided in Chapter 3.

Chapter 5

Future research ideas in group testing

During the progress of this dissertation, I have come up with some new ideas that could be interesting and would contribute significantly to the statistical literature in group testing. I herein describe few research problems that I did not have time to explore but I plan to pursue in future.

5.1 Group testing for multiple infections

- In Chapter 2, we modeled the group testing data for multiple infections from the IPP two-stage hierarchical algorithm. An immediate extension of this work could be to incorporate test results from more than two stages; an example of such group testing method is the halving algorithm. We found on a separate piece of work that testing performed in more than two stages can actually increase individuals' screening efficiency. Consequently, this extension will enable one to estimate disease prevalences using data from more cost efficient testing protocols. I also plan to investigate optimal pool sizes for the multiple-stage pooling algorithm which involves multiple correlated infections. In Chapter 2, we used the pool size, suggested by Tebbs, McMahan, and Bilder (2013), which maximizes individuals' screening efficiency. However, when the goal is estimation, it is natural to use the optimal pool size which results in the most precise estimates.

- A further advancement of the problem discussed above is possible by adjusting for individual-level covariate information. Zhang, Bilder, and Tebbs (2013) first proposed a likelihood-based inference for the IPP two-stage algorithm. A future research can extend this work to more stages in both frequentist's and Bayesian framework.

- The IPP pooling algorithm in Chapter 2 assumes the availability of a discriminatory assay for both master pool testing in Stage 1 and individuals' retesting in Stage 2. A future work can generalize our Bayesian model in Chapter 2 to allow for other types of assays that are commonly used in group testing. For example, a variant of the IPP pool testing can use a discriminatory assay in Stage 1 and a disease specific assay in Stage 2. This provides added flexibility in testing and is particularly useful when more accurate diagnoses are desired to retest individuals from positive pools. Another such testing algorithm can apply a non-discriminatory assay in Stage 1 and a disease specific assay in Stage 2. In my future research, I plan to survey the group testing protocols implemented by American Red Cross, German Red Cross, etc., and then develop statistical methods for those applications.

5.2 Group testing for single infection

- In Chapter 3, we proposed a regression method that accounts for dilution effects through a parametric submodel. Even though the parametric approach is flexible, a poor choice of the submodel can seriously compromise estimation precision. An alternative approach could be to estimate the submodel nonparametrically. Particularly, one can find order-restricted maximum likelihood estimators of the submodel by exploiting the property that an assay's pool testing sensitivity is increasing as a function of the number of true positives. This semiparametric approach (i.e., parametric primary regression model and non-

parametric submodel) will be appropriate even with a reasonably large sample size. This project is in progress; we are currently in the exploratory stage.

- The dilution model in Chapter 3 requires a large sample for its validity. Even though such a requirement is not prohibitive, a Bayesian approach can be more suitable and can offer several advantages. First, carefully constructed prior distributions can substantially reduce the large sample requirements. The priors can be easily elicited using historical data and assay product literature information. Second, one can relax the assumption that individual testing sensitivity and specificity (S_e and S_p) are known constants. Inference is possible conventionally by Markov Chain Monte Carlo techniques, such as Gibbs sampler and Metropolis-Hastings algorithm.

- A more general version of the dilution method in Chapter 3 can include random effects to account for regional variability. To accomplish this, one can combine our method in Chapter 3 and the random effects model in Chen, Tebbs, and Bilder (2009). Hypothesis tests for dilution effects and for regional variability can be done as before. A general regression framework could be of practical interest. This will be a challenging problem because of the complex model structure for group testing and also because of additional computational burden by including random effects.

5.3 Group testing coupled with measurement error in covariates

- I plan to study the asymptotic properties of naive maximum likelihood estimators in the presence of measurement error in covariates. The naive regression estimates for individual testing are usually attenuated; similar results are expected for the naive estimates using group testing. However, previous robustness studies (Hung and Swallow, 1999; Huang and Tebbs, 2009) suggest

that inference using group testing should be less affected by the measurement error. If such a robustness outcome is revealed, we might be able to develop a formal hypothesis test which detects measurement errors. Furthermore, this outcome can provide several other appealing features. First, practitioners will be more encouraged to use group testing as an alternative to individual testing, whereby substantial cost saving is possible. Second, the hypothesis test does not require replicate measurements, unlike the hypothesis test for individual testing. Note that replicate measurements can often be prohibitively costly, both, economically and logistically.

- Another measurement error project, which can be supremely interesting, is the generalization of the individual testing regression models in Tsiatis and Ma (2004) to allow for group testing. The authors proposed functional models which treat error-prone covariates as random variables that follow an unknown parametric distribution. This study is limited to the generalized linear models and is not applicable for more sophisticated data structures, such as group testing where regression models usually fall outside the range of generalized linear models.

Bibliography

S. Allwright, F. Bradley, J. Long, J. Barry, L. Thornton, and J. Parry. Prevalence of antibodies to hepatitis b, hepatitis c, and hiv and risk factors in irish prisoners: results of a national cross sectional survey. *British Medical Journal*, 321:78–82, 2000.

C. Bilder and J. Tebbs. Empirical Bayesian estimation of the disease transmisssion probability in multiple-vector-transfer designs. *Biometrical Journal*, 47:502–516, 2005.

C. Bilder and J. Tebbs. Bias, efficiency, and agreement for group-testing regression models. *Journal of Statistical Computation and Simulation*, 79:67–80, 2009.

C. Bilder, J. Tebbs, and P. Chen. Informative retesting. *Journal of the American Statistical Association*, 105:942–955, 2010.

B. Branson and J. Mermin. Establishing the diagnosis of HIV infection: New tests and a new algorithm for the United States. *Journal of Clinical Virology*, 52:S3–4, 2011.

P. Burrows. Improved estimation of pathogen transmission rates by group testing. *Phytopathology*, 77:363–365, 1987.

M. Busch, S. Caglioti, E. Robertson, J. McAuley, L. Tobler, H. Kamel, J. Linnen, V. Shyamala, P. Tomasulo, and S. Kleinman. Screening the blood supply for West Nile virus RNA by nucleic acid amplification testing. *New England Journal of Medicine*, 353:460–467, 2005.

R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective*. Boka Raton, Florida: Chapman and Hall/CRC, 2006.

CDC. Recommendations for the Laboratory-Based Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Morbidity and Mortality Weekly Report*, 2015. URL `http://www.cdc.gov/mmwr`.

Y. Chaubey and W. Li. Comparison between maximum likelihood and bayes methods for estimation of binomial probability with samples compositing. *Journal of Official Statistics*, 11:379–390, 1995.

C. Chen and W. Swallow. Using group testing to estimate a proportion, and to test the binomial model. *Biometrics*, 46:1035–1046, 1990.

P. Chen, J. Tebbs, and C. Bilder. Group testing regression models with fixed and random effects. *Biometrics*, 65:1270–1278, 2009.

X. Chi, X. Lou, M. Yang, and Q. Shu. An optimal DNA pooling strategy for progressive fine mapping. *Genetica*, 135:267–281, 2009.

A. Delaigle and P. Hall. Nonparametric regression with homogeneous group testing data. *The Annals of Statistics*, 2012.

A. Delaigle and P Hall. Nonparametric methods for group testing data, taking dilution into account. *Biometrika*, 102:871–887, 2015.

A. Delaigle and A. Meister. Nonparametric regression analysis for group testing data. *Journal of the American Statistical Association*, 106:640–650, 2011.

A. Delaigle and W. Zhou. Nonparametric and parametric estimators of prevalence from group testing data with aggregated covariates. *Journal of the American Statistical Association*, 110:1785–1796, 2015.

N. Dhand, W. Johnson, and J. Toribio. A Bayesian approach to estimate OJD prevalence from pooled fecal samples of variable pool size. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:452–473, 2010.

R. Dorfman. The detection of defective members of large populations. *Annals of Mathematical Statistics*, 14:436–440, 1943.

J. Fahey, P. Ourisson, and F. Degnan. Pathogen detection, testing, and control in fresh broccoli sprouts. *Nutrition Journal*, 5:13, 2006.

C. Farrington. Estimating prevalence by group testing using generalized linear models. *Statistics in Medicine*, 11:1591–1597, 1992.

J. L. Gastwirth and W. O. Johnson. Screening with cost-effective quality control: Potential applications to HIV and drug testing. *Journal of the American Statistical Association*, 89:972–981, 1994.

C. Gaydos, C. Cartwright, P. Colaninno, J. Welsch, J. Holden, S. Yo, E. Webb, C. Anderson, R. Bertuzis, L. Zhang, T. Miller, G. Leckie, K. Abravaya, and J. Robinson. Performance of the Abbott RealTime CT/NG for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Journal of Clinical Microbiology*, 48: 3236–3243, 2010.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

K. Hanson, B. Alexander, C. Woods, C. Petti, and L. Reller. Validation of laboratory screening criteria for herpes simplex virus testing of cerebrospinal fluid. *Journal of Clinical Microbiology*, 45:721–724, 2007.

T. Hanson, W. Johnson, and J. Gastwirth. Bayesian inference for prevalence and diagnostic test accuracy based on dual-pooled screening. *Biostatistics*, 7:41–57, 2006.

W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

M. Hourfar, C. Jork, V. Schottstedt, M. Weber-Schehl, V. Brixner, M. Busch, G. Geusendam, K. Gubbe, C. Mahnhardt, W. Mayr-Wohlfar, L. Pichl, W. Roth, M. Schmidt, E. Seifried, and D. Wright. Experience of German Red Cross blood donor services with nucleic acid testing: Results of screening more than 30 million blood donations for human immunodeficiency virus, hepatitis C virus, and hepatitis B virus. *Transfusion*, 48:1558–1566, 2008.

X. Huang. An improved test of latent-variable model misspecification in structural measurement error models for group testing data. *Statistics in Medicine*, 28: 3316–3327, 2009.

X. Huang and J. Tebbs. On latent-variable model misspecification in structural measurement error models for binary response. *Biometrics*, 65:710–718, 2009.

M. Hudgens and M. Halloran. Towards causal inference with interference. *Journal of the American Statistical Association*, 103:832–842, 2008.

J. Hughes-Oliver and W. Rosenberger. Efficient estimation of the prevalence of multiple rare traits. *Biometrika*, 87:315–327, 2000.

M. Hung and W. Swallow. Robustness of group testing in the estimation of proportions. *Biometrics*, 55:231–237, 1999.

F. Hwang. An optimum nested procedure in binomial group testing. *Journal of the American Statistical Association*, 32:939–943, 1976.

J. Ibrahim, M. Chen, Y. Gwon, and F. Chen. The power prior: Theory and applications. *Statistics in Medicine*, 34:3724–3749, 2015.

S. Jirsa. Pooling specimens: A decade of successful cost savings. *National STD Prevention Conference*, 2008.

W. Johnson and L. Pearson. Dual screening. *Biometrics*, 55:867–873, 1999.

Inc./Denver JSI Research & Training Institute. The future of Infertility Prevention Project Health Impact Assessment: Policy Implications and Recommendations in Light of Passage of the Patient Protection and Affordable Care Act, July 25, 2012. *http://www.jsi.com*, 2015.

H. Kim, M. Hudgens, J. Dreyfuss, D. Westreich, and C. Pilcher. Comparison of group testing algorithms for case identification in the presence of testing error. *Biometrics*, 63:1152–1163, 2007.

M. Krajden, D. Cook, A. Mak, K. Chu, N. Chahil, M. Steinberg, M. Rekart, and M. Gilbert. Pooled nucleic acid testing increases the diagnostic yield of acute

HIV infections in a high-risk population compared to 3rd and 4th generation HIV enzyme immunoassays. *Journal of Clinical Virology*, 61:132–137, 2014.

J. Lewis, V. Lockary, and S. Kobic. Cost savings and increased efficiency using a stratified specimen pooling strategy for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *Sexually Transmitted Diseases*, 39:46–48, 2012.

C. Lindan, M. Mathur, S. Kumta, H. Jerajani, A. Gogate, J. Schachter, and J. Moncada. Utility of pooled urine specimens for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in men attending public sexually transmitted infection clinics in Mumbai, India, by PCR. *Journal of Clinical Microbiology*, 43: 1674–1677, 2005.

E. Litvak, X. Tu, and M. Pagano. Screening for the presence of a disease by pooling sera samples. *Journal of the American Statistical Association*, 89:424–434, 1994.

A. Liu, C. Liu, Z. Zhang, and P. Albert. Optimality of group testing in the presence of misclassification. *Biometrika*, 99:245–251, 2012.

T. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B (Methodology)*, 44:226–233, 1982.

C. McMahan, J. Tebbs, and C. Bilder. Two-dimensional informative array testing. *Biometrics*, 68:793–804, 2012a.

C. McMahan, J. Tebbs, and C. Bilder. Informative Dorfman screening. *Biometrics*, 68:287–296, 2012b.

C. McMahan, J. Tebbs, and C. Bilder. Regression models for group testing data with pool dilution effects. *Biostatistics*, 14:284–298, 2013.

J. Mendoza-Blanco, X. Tu, and S. Iyengar. Bayesian inference on prevalence using a missing-data approach with simulation-based techniques: Applications to HIV screening. *Statistics in Medicine*, 15:2161–2176, 1996.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *American Institute of Physics*, 21: 1087, 1953.

G. Molenberghs and G. Verbeke. Likelihood ratio, score, and wald tests in a constrained parameter space. *The American Statistician*, 61:22–27, 2007.

S. O'Brien, Q. Yi, W. Fan, V. Scalia, M. Fearon, and J. Allain. Current incidence and residual risk of HIV, HBV and HCV at Canadian Blood Services. *Vox Sanguinis*, 103:83–86, 2012.

R. M. Phatarfod and A. Sudbury. The use of a square array scheme in blood testing. *Statistics in Medicine*, 13:2337–2343, 1994.

C. Pilcher, S. Fiscus, T. Nguyen, E. Foust, L. Wolf, D. Williams, R. Ashby, and J. O'Dowd. Detection of acute infections during HIV testing in North Carolina. *New England Journal of Medicine*, 352:1873–1883, 2005.

K. Remlinger, J. Hughes-Oliver, S. Young, and R. Lam. Statistical design of pools using optimal coverage and minimal collision. *Technometrics*, 48:133–143, 2006.

M. Schmidt, L. Pichl, C. Jork, M. Hourfar, V. Schottstedt, F. Wagner, E. Seifried, T. Müller, J. Bux, and J. Saldanha. Blood donor screening with cobas s 201/cobas TaqScreen MPX under routine conditions at German Red Cross Institutes. *Vox Sanguinis*, 98:37–46, 2010.

S. Self and K. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82:605–610, 1987.

M. Sobel and R. Elashoff. Group testing with a new goal, estimation. *Biometrika*, 62:181–193, 1975.

A. Sterrett. On the detection of defective members of large populations. *Annals of Mathematical Statistics*, 28:1033–1036, 1957.

S. Stramer, U. Wend, D. Candotti, G. Foster, F. Hollinger, R. Dodd, J. Allain, and W. Gerlich. Nucleic acid testing to detect HBV infection in blood donors. *New England Journal of Medicine*, 364:236–247, 2011.

S. Stramer, E. Notari, D. Krysztof, and R. Dodd. Hepatitis B virus testing by minipool nucleic acid testing: Does it improve blood safety? *Transfusion*, 53: 2449–2458, 2013.

J. Tebbs, C. McMahan, and C. Bilder. Two-stage hierarchical group testing for multiple infections with application to the Infertility Prevention Project. *Biometrics*, 69:1064–1073, 2013.

K. Thompson. Estimation of the proportion of vectors in a natural population of insects. *Biometrics*, 18:568–578, 1962.

A. Tsiatis and Y. Ma. Locally efficient semiparametric estimators for functional measurement error models. *Biometrika*, 91:835–848, 2004.

X.M. Tu, E. Litvak, and M. Pagano. On the informativeness and accuracy of pooled testing when estimating prevalence of a rare disease: Application to HIV screening. *Biometrika*, 82:287–298, 1995.

T. Van, J. Miller, D. Warshauer, E. Reisdorf, D. Jerrigan, R. Humes, and P. Shult. Pooling nasopharyngeal/throat swab speciments to increase testing capacity for influenza viruses by PCR. *Journal of Clinical Microbiology*, 50:891–896, 2012.

S. Vansteelandt, E. Goetghebeur, and T. Verstraeten. Regression models for disease prevalence with diagnostic tests on pools of serum samples. *Biometrics*, 56:1126–1133, 2000.

M. A. Wahed, D. Chowdhury, B. Nermell, S. I. Khan, M. Ilias, M. Rahman, L. A. Persson, and M. Vahter. A modified routine analysis of arsenic content in drinking water in Bangladesh by hydride generation-atomic absorption spectrophotometry. *J. Health, Population and Nutrition*, 24:36–41, 2006.

B. Wang, S. Han, C. Cho, J. Han, Y. Cheng, S. Lee, G. Galappaththy, K. Thimasarn, M. Soe, H. Oo, M. Kyaw, and E. Han. Comparison of microscopy, nested PCR, and real-time PCR assays using high-throughput screening of pooled samples for diagnosis of malaria in asymptomatic carriers from areas of endemicity in Myanmar. *Journal of Clinical Microbiology*, 52:1838–1845, 2014a.

D. Wang, C. McMahan, C. Gallagher, and K. Kulasekera. Semiparametric group testing regression models. *Biometrika*, 101:587–598, 2014b.

D. Wang, C. McMahan, and C. Gallagher. A general parametric regression framework for group testing data with dilution effects. *Statistics in Medicine*, 34:3606–3621, 2015.

Y. Wang. On the number of successes in independent trials. *Statistica Sinica*, 3: 295–312, 1993.

L. Wein and S. Zenios. Pooled testing for HIV screening: Capturing the dilution effect. *Operations Research*, 44:543–569, 1996.

M. Xie. Regression analysis of group testing samples. *Statistics in Medicine*, 20: 1957–1969, 2001.

S. Zenios and L. Wein. Pooled testing for HIV prevalence estimation: Exploiting the dilution effect. *Statistic in Medicine*, 17:1447–1467, 1998.

B. Zhang, C. Bilder, and J. Tebbs. Group testing regression model estimation when case identification is a goal. *Biometrical Journal*, 55:173–189, 2013.

Z. Zhang, C. Liu, S. Kim, and A. Liu. Prevalence estimation subject to misclassification: The mis-substitution bias and some remedies. *Statistics in Medicine*, 33: 4482–4500, 2014.

Appendix A

Chapter 2 supplementary materials

A.1 Generalization of estimation methods to include $J \geq 2$ infections.

It is straightforward to generalize our estimation procedure in Section 2.3 to include $J \geq 2$ infections. We continue to assume that a single discriminating assay is used in both stages of the algorithm below. For potential applications, see Section 2.6.

POOLING ALGORITHM

Stage 1: Individuals are randomly assigned to master pools. Each pool is tested for each of J infections using a single assay. A single assay detects all infections simultaneously.

Stage 2: Individuals in pools that

- test negatively for all J infections are diagnosed as negative for all infections.
- test positively for at least one infection are retested (individually) for all J infections using the same assay in Stage 1. Diagnoses for all infections are made from the outcomes of the individual tests.

Suppose N individuals are to be tested using the two-stage algorithm above for $J \geq 2$ infections. Let $\widetilde{\mathbf{Y}}_{ik} = (\widetilde{Y}_{i1k}, \widetilde{Y}_{i2k}, ..., \widetilde{Y}_{iJk})'$ denote the vector of true statuses for the ith individual in the kth pool, for $i = 1, 2, ..., c_k$ and $k = 1, 2, ..., K$, where $N = \sum_{k=1}^{K} c_k$. Let $\widetilde{\mathbf{Z}}_k = (\widetilde{Z}_{1k}, \widetilde{Z}_{2k}, ..., \widetilde{Z}_{Jk})'$ and $\mathbf{Z}_k = (Z_{1k}, Z_{2k}, ..., Z_{Jk})'$ denote the vector of true and observed statuses for the kth master pool, respectively. If the kth master pool tests positively for at least one infection in Stage 1, the testing response

vector for ith individual in Stage 2 is $\mathbf{Y}_{ik} = (Y_{i1k}, Y_{i2k}, ..., Y_{iJk})'$. Denote the $2J \times 1$ vector of assay accuracies by $\boldsymbol{\delta} = (S_{e:1}, S_{e:2}, ..., S_{e:J}, S_{p:1}, S_{p:2}, ..., S_{p:J})'$.

We use the notation adopted by Hughes-Oliver and Rosenberger (2000), referenced in the manuscript. Let $\omega = (\omega_1\omega_2 \ldots \omega_J)$ denote the J-tuple where $\omega_j \in \{0, 1\}$, for $j = 1, 2, ..., J$, and let Ω denote the collection of the 2^J ω outcomes. The probability of the outcome ω is denoted by $p_\omega \in (0, 1)$ so that $\sum_{\omega\in\Omega} p_\omega = 1$. Let $\mathbf{p}$ denote the $2^J \times 1$ vector consisting of $\{p_\omega : \omega \in \Omega\}$. The probability mass function (pmf) of $\widetilde{\mathbf{Y}}_{ik}$ is

$$\text{pr}(\widetilde{\mathbf{Y}}_{ik} = \widetilde{\mathbf{y}}|\mathbf{p}) = \prod_{\omega\in\Omega} p_\omega^{\widetilde{v}_\omega}, \tag{A.1}$$

where $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_J)'$ and $\widetilde{v}_\omega = \prod_{j=1}^J \widetilde{y}_j^{\omega_j}(1 - \widetilde{y}_j)^{1-\omega_j}$. For example, when $J = 2$, we have $\Omega = \{00, 10, 01, 11\}$, $\mathbf{p} = (p_{00}, p_{10}, p_{01}, p_{11})'$, and

$$\begin{aligned}
\widetilde{v}_{00} &= (1 - \widetilde{y}_1)(1 - \widetilde{y}_2) \\
\widetilde{v}_{10} &= \widetilde{y}_1(1 - \widetilde{y}_2) \\
\widetilde{v}_{01} &= (1 - \widetilde{y}_1)\widetilde{y}_2 \\
\widetilde{v}_{11} &= \widetilde{y}_1\widetilde{y}_2.
\end{aligned}$$

In this case, the pmf in Equation (A.1) reduces to the pmf in Section 2.3.

Let $\boldsymbol{\theta} = (\mathbf{p}', \boldsymbol{\delta}')'$. Under the same assumptions described in Section 2.3 and Section 2.6 (i.e., non-differential assay error, conditional independence of testing results given the true statuses, no-interference among diseases with regard to $S_{e:j}$ and $S_{p:j}$), the joint distribution of the observed data $\{\mathbf{Z}, \mathbf{Y}\}$ and the latent data $\widetilde{\mathbf{Y}}$, conditional on $\boldsymbol{\theta}$, is given by

$$\begin{aligned}
\pi(\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}}|\boldsymbol{\theta}) = &\prod_{k=1}^K \prod_{i=1}^{c_k} \prod_{\omega\in\Omega} p_\omega^{\widetilde{V}_{(\omega)ik}} \\
&\times \left[\prod_{j=1}^J \prod_{k=1}^K \left(S_{e:j}^{Z_{jk}} \overline{S}_{e:j}^{1-Z_{jk}} \right)^{I(\sum_{i=1}^{c_k} \widetilde{Y}_{ijk} > 0)} \left(S_{p:j}^{1-Z_{jk}} \overline{S}_{p:j}^{Z_{jk}} \right)^{I(\sum_{i=1}^{c_k} \widetilde{Y}_{ijk} = 0)} \right. \\
&\left. \times \left\{ \prod_{i=1}^{c_k} S_{e:j}^{Y_{ijk}\widetilde{Y}_{ijk}} \overline{S}_{e:j}^{(1-Y_{ijk})\widetilde{Y}_{ijk}} S_{p:j}^{(1-Y_{ijk})(1-\widetilde{Y}_{ijk})} \overline{S}_{p:j}^{Y_{ijk}(1-\widetilde{Y}_{ijk})} \right\}^{I(\sum_{j'=1}^J Z_{j'k} > 0)} \right],
\end{aligned}$$

where $\widetilde{V}_{(\omega)ik} = \prod_{j=1}^{J} \widetilde{Y}_{ijk}^{\omega_j}(1-\widetilde{Y}_{ijk})^{1-\omega_j}$, $\overline{S}_{e:j} = 1 - S_{e:j}$, and $\overline{S}_{p:j} = 1 - S_{p:j}$. Our goal is to estimate the $2^J + 2J$ parameters in $\boldsymbol{\theta}$.

As in Section 2.3, we elicit independent beta prior distributions for the assay test accuracies; i.e., $S_{e:j} \sim \text{beta}(a_{S_{e:j}}, b_{S_{e:j}})$ and $S_{p:j} \sim \text{beta}(a_{S_{p:j}}, b_{S_{p:j}})$, for $j = 1, 2, ..., J$, where all hyperparameters are known. We specify a Dirichlet prior distribution for $\mathbf{p}$; specifically,

$$\mathbf{p} \sim \pi(\mathbf{p}) \propto \prod_{\omega \in \Omega} p_{\omega}^{\alpha_\omega - 1},$$

and derive full conditional distributions. For the assay accuracies, these distributions are $S_{e:j}|\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}} \sim \text{beta}(a^*_{S_{e:j}}, b^*_{S_{e:j}})$ and $S_{p:j}|\mathbf{Z}, \mathbf{Y}, \widetilde{\mathbf{Y}} \sim \text{beta}(a^*_{S_{p:j}}, b^*_{S_{p:j}})$, for $j = 1, 2, ..., J$, where

$$\begin{aligned}
a^*_{S_{e:j}} &= a_{S_{e:j}} + \sum_{k=1}^{K} \left\{ Z_{jk}\widetilde{Z}_{jk} + I\left(\sum_{j'=1}^{J} Z_{j'k} > 0\right) \sum_{i=1}^{c_k} Y_{ijk}\widetilde{Y}_{ijk} \right\} \\
b^*_{S_{e:j}} &= b_{S_{e:j}} + \sum_{k=1}^{K} \left\{ (1-Z_{jk})\widetilde{Z}_{jk} + I\left(\sum_{j'=1}^{J} Z_{j'k} > 0\right) \sum_{i=1}^{c_k} (1-Y_{ijk})\widetilde{Y}_{ijk} \right\} \\
a^*_{S_{p:j}} &= a_{S_{p:j}} + \sum_{k=1}^{K} \left\{ (1-Z_{jk})(1-\widetilde{Z}_{jk}) + I\left(\sum_{j'=1}^{J} Z_{j'k} > 0\right) \sum_{i=1}^{c_k} (1-Y_{ijk})(1-\widetilde{Y}_{ijk}) \right\} \\
b^*_{S_{p:j}} &= b_{S_{p:j}} + \sum_{k=1}^{K} \left\{ Z_{jk}(1-\widetilde{Z}_{jk}) + I\left(\sum_{j'=1}^{J} Z_{j'k} > 0\right) \sum_{i=1}^{c_k} Y_{ijk}(1-\widetilde{Y}_{ijk}) \right\}
\end{aligned}$$

and $\widetilde{Z}_{jk} = I(\sum_{i=1}^{c_k} \widetilde{Y}_{ijk} > 0)$. For the prevalence parameter $\mathbf{p}$, the full conditional distribution is Dirichlet; i.e., $\mathbf{p}|\widetilde{\mathbf{Y}} \sim \text{Dirichlet}(\boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is a $2^J \times 1$ vector of the elements of $\{\alpha_\omega + \sum_{k=1}^{K}\sum_{i=1}^{c_k} \widetilde{V}_{(\omega)ik} : \omega \in \Omega\}$, where $\widetilde{V}_{(\omega)ik} = \prod_{j=1}^{J} \widetilde{Y}_{ijk}^{\omega_j}(1-\widetilde{Y}_{ijk})^{1-\omega_j}$, for $\omega_j \in \{0, 1\}$.

Let $\widetilde{\mathbf{V}}_{ik}$ denote the $2^J \times 1$ vector of the elements of $\{\widetilde{V}_{(\omega)ik} : \omega \in \Omega\}$. The conditional distribution of $\widetilde{\mathbf{V}}_{ik}$ given $\{\widetilde{\mathbf{Y}}_{k(i)}, \mathbf{p}, \boldsymbol{\delta}, \mathbf{Y}, \mathbf{Z}\}$ is multinomial with cell probabilities $\zeta^{ik}_{\omega}/\zeta^{ik}$, where

$$\begin{aligned}
\zeta^{ik}_{\omega} = p_\omega \prod_{j=1}^{J} &\left(S_{e:j}^{Z_{jk}} \bar{S}_{e:j}^{1-Z_{jk}}\right)^{\omega_j} \left\{ \left(S_{e:j}^{Z_{jk}} \bar{S}_{e:j}^{1-Z_{jk}}\right)^{\gamma_{ijk}} \left(S_{p:j}^{1-Z_{jk}} \bar{S}_{p:j}^{Z_{jk}}\right)^{1-\gamma_{ijk}} \right\}^{1-\omega_j} \\
&\times \left\{ \left(S_{e:j}^{Y_{ijk}} \bar{S}_{e:j}^{1-Y_{ijk}}\right)^{I(\sum_{j'=1}^{J} Z_{j'k} > 0)} \right\}^{\omega_j} \left\{ \left(S_{p:j}^{1-Y_{ijk}} \bar{S}_{p:j}^{Y_{ijk}}\right)^{I(\sum_{j'=1}^{J} Z_{j'k} > 0)} \right\}^{1-\omega_j},
\end{aligned}$$

$\zeta^{ik} = \sum_{\omega' \in \Omega} \zeta^{ik}_{\omega'}$, and $\gamma_{ijk} = I(\sum_{i' \neq i} \widetilde{Y}_{i'jk} > 0)$. When $J = 2$, the general expression for ζ^{ik}_{ω} admits the four multinomial cell probabilities stated in Section 2.3. Once $\widetilde{\mathbf{V}}_{ik}$ is sampled from its conditional distribution, $\widetilde{\mathbf{Y}}_{ik}$ can be uniquely determined as shown in Section 2.3 when $J = 2$. With the conditional distributions of $\mathbf{p}$, $\boldsymbol{\delta}$, and $\widetilde{\mathbf{V}}_{ik}$ available, one can sample from these distributions using the Gibbs sampler we described in Section 2.3.

A.2 Complete simulation results from Section 2.4.

This appendix contains the complete set of simulation results from Section 2.4. The following figures are provided:

- Figure A.1: Prevalence estimates with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$
- Figure A.2: Assay accuracy estimates with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$
  - flat priors for $S_{e:j}$ and $S_{p:j}$
- Figure A.3: Prevalence estimates with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$
- Figure A.4: Assay accuracy estimates with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$
  - informative priors for $S_{e:j}$ and $S_{p:j}$
- Figure A.5: Prevalence estimates with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$
- Figure A.6: Assay accuracy estimates with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$
  - flat priors for $S_{e:j}$ and $S_{p:j}$
- Figure A.7: Prevalence estimates with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$
- Figure A.8: Assay accuracy estimates with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$
  - informative priors for $S_{e:j}$ and $S_{p:j}$

Note that Figures A.1 and A.2 are the same as Figures 2.1 and 2.2 in Chapter 2.

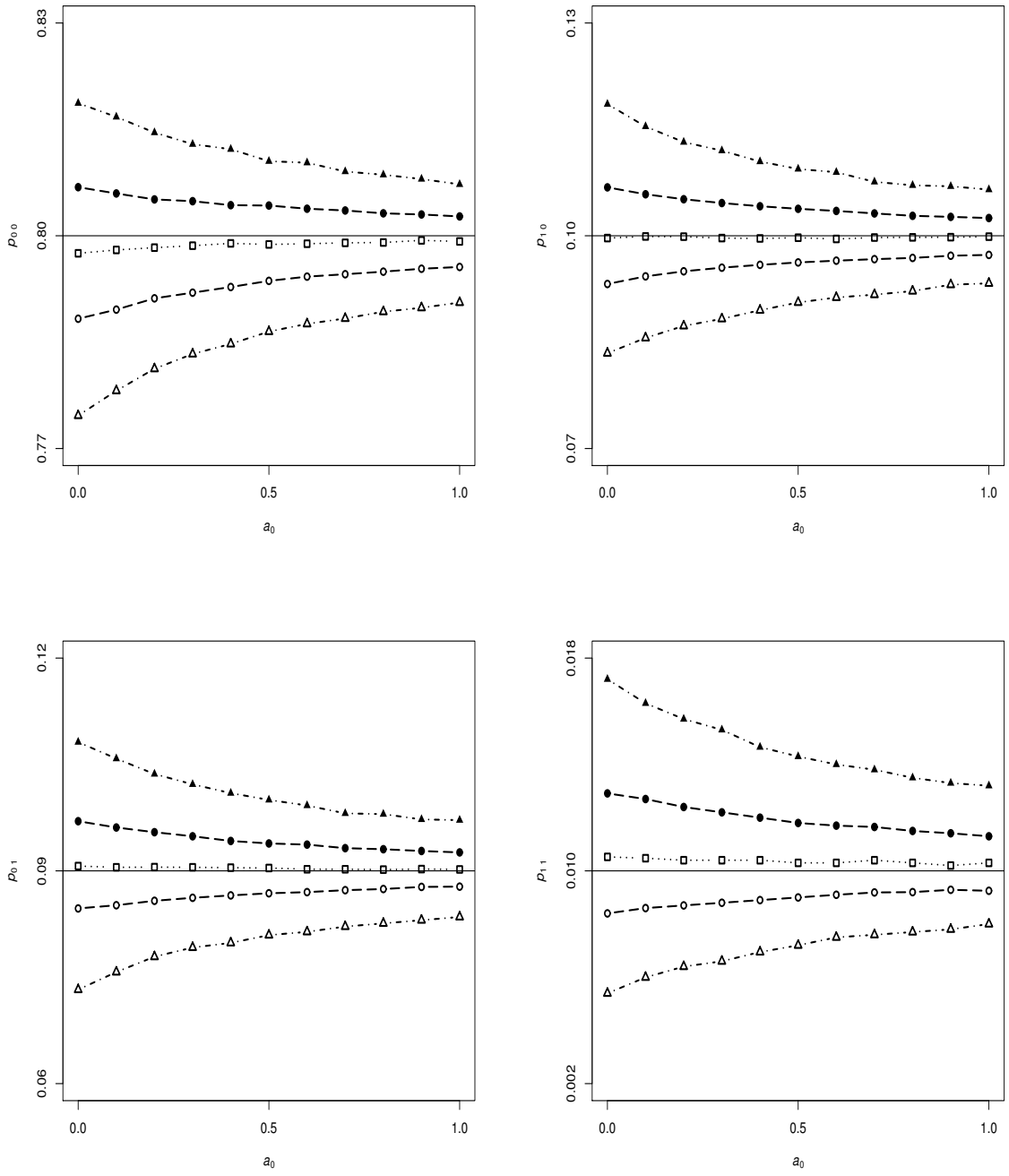

Figure A.1: Simulation results with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$ are provided. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

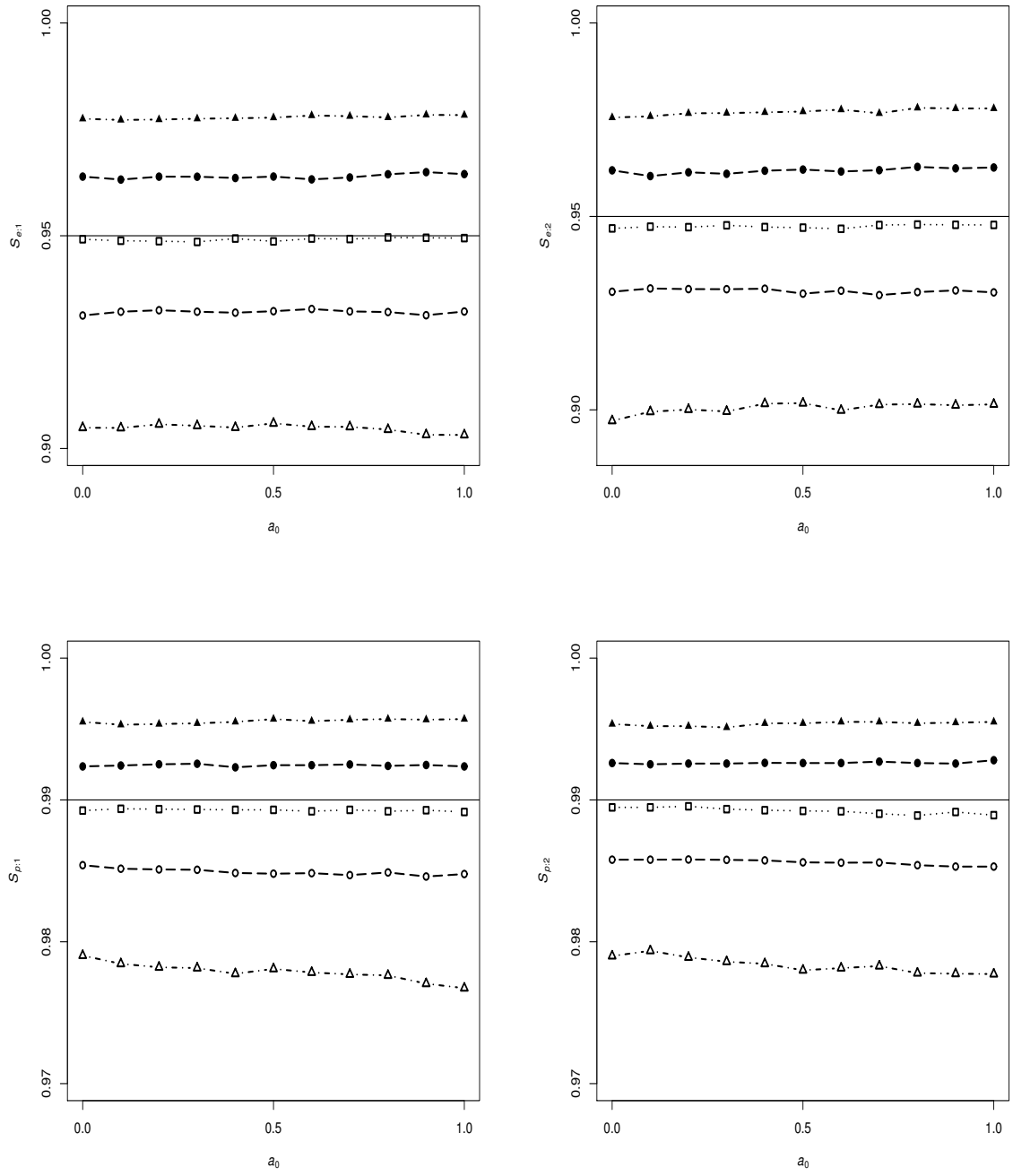

Figure A.2: Simulation results with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $S_{e:j}$ and $S_{p:j}$ are provided. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

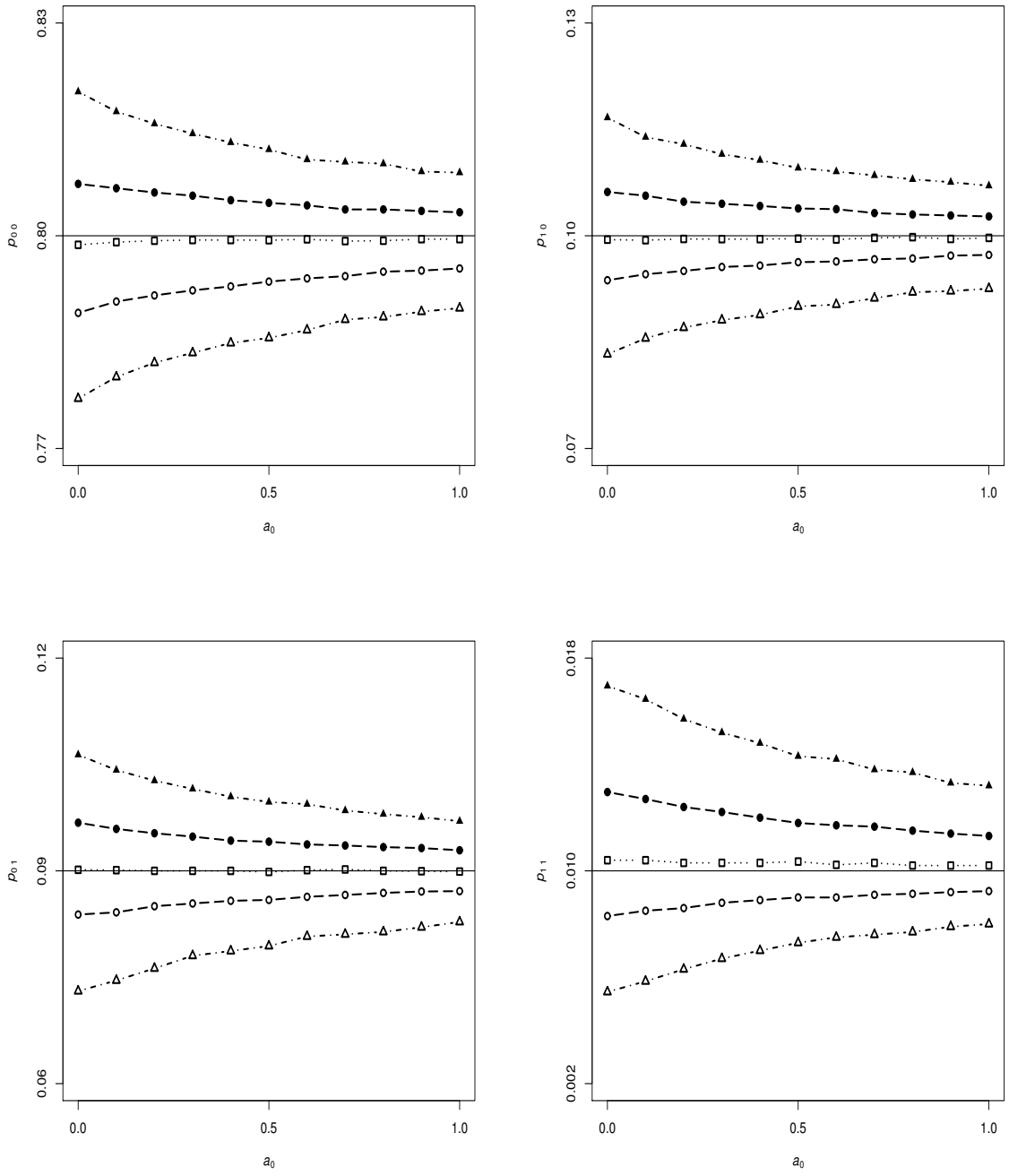

Figure A.3: Simulation results with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$ are provided. Informative priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(109.0, 6.7)$ and $S_{p:j} \sim \text{beta}(55.2, 1.6)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

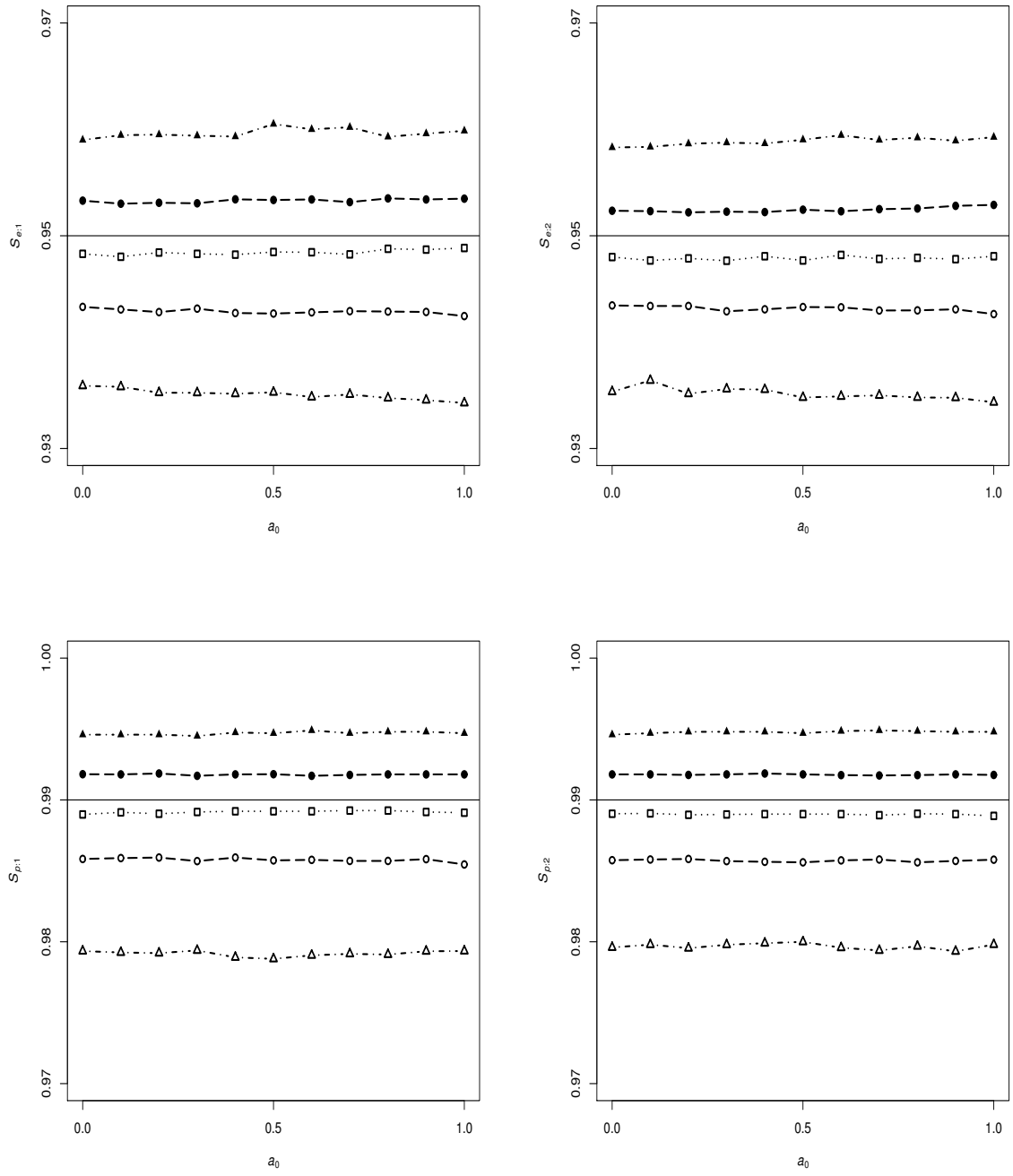

Figure A.4: Simulation results with $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $S_{e:j}$ and $S_{p:j}$ are provided. Informative priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(109.0, 6.7)$ and $S_{p:j} \sim \text{beta}(55.2, 1.6)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

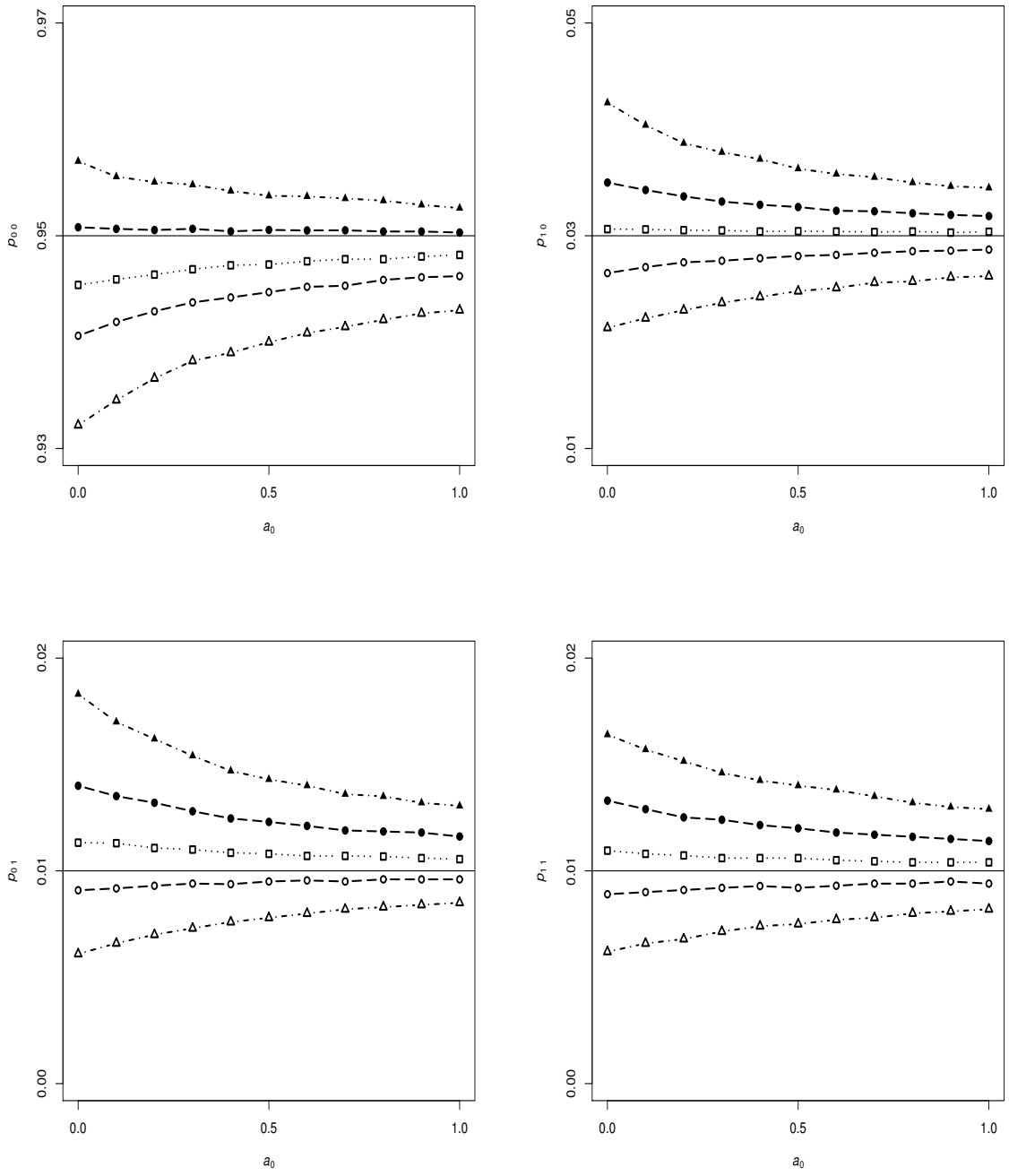

Figure A.5: Simulation results with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$ are provided. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

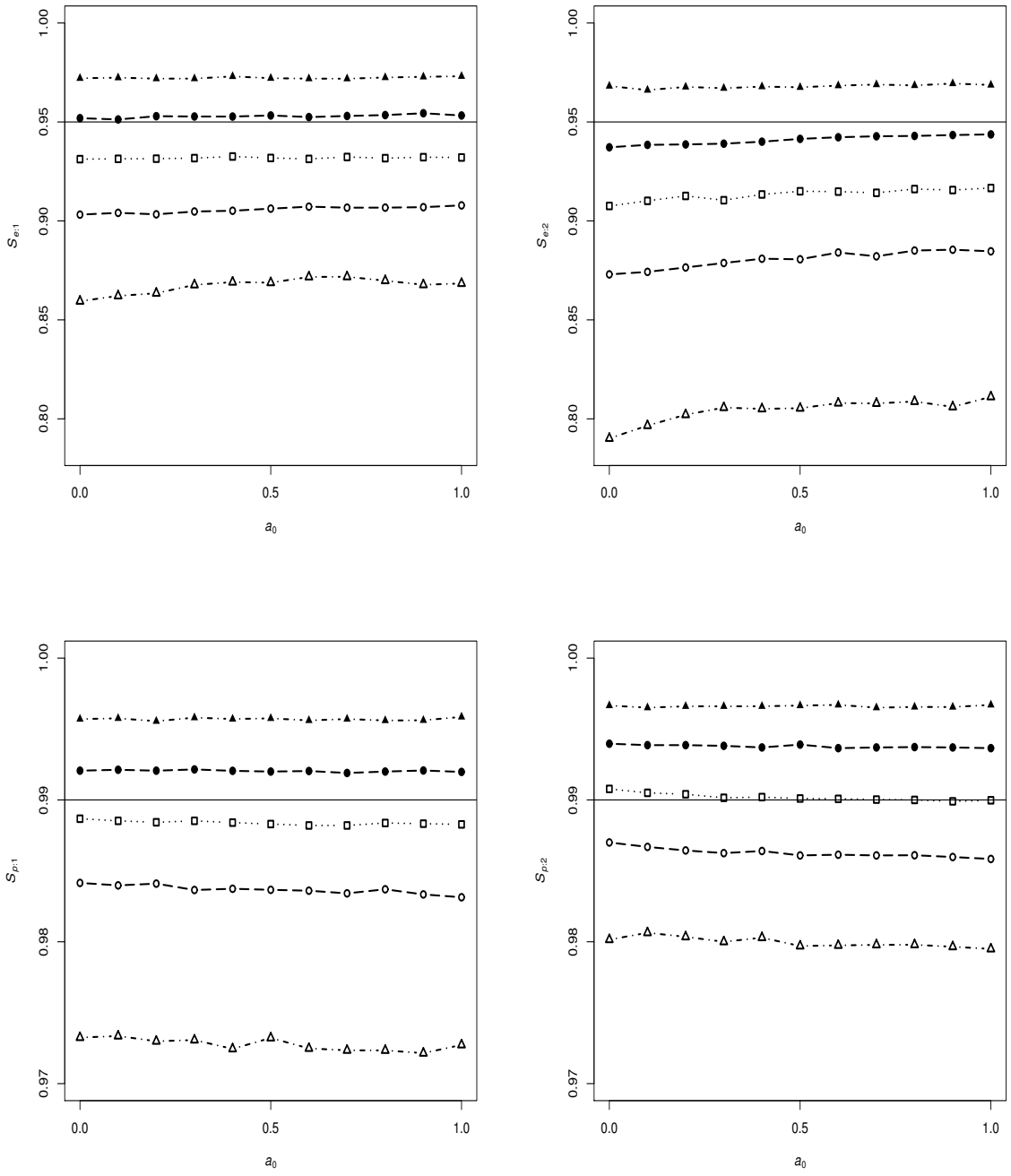

Figure A.6: Simulation results with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $S_{e:j}$ and $S_{p:j}$ are provided. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

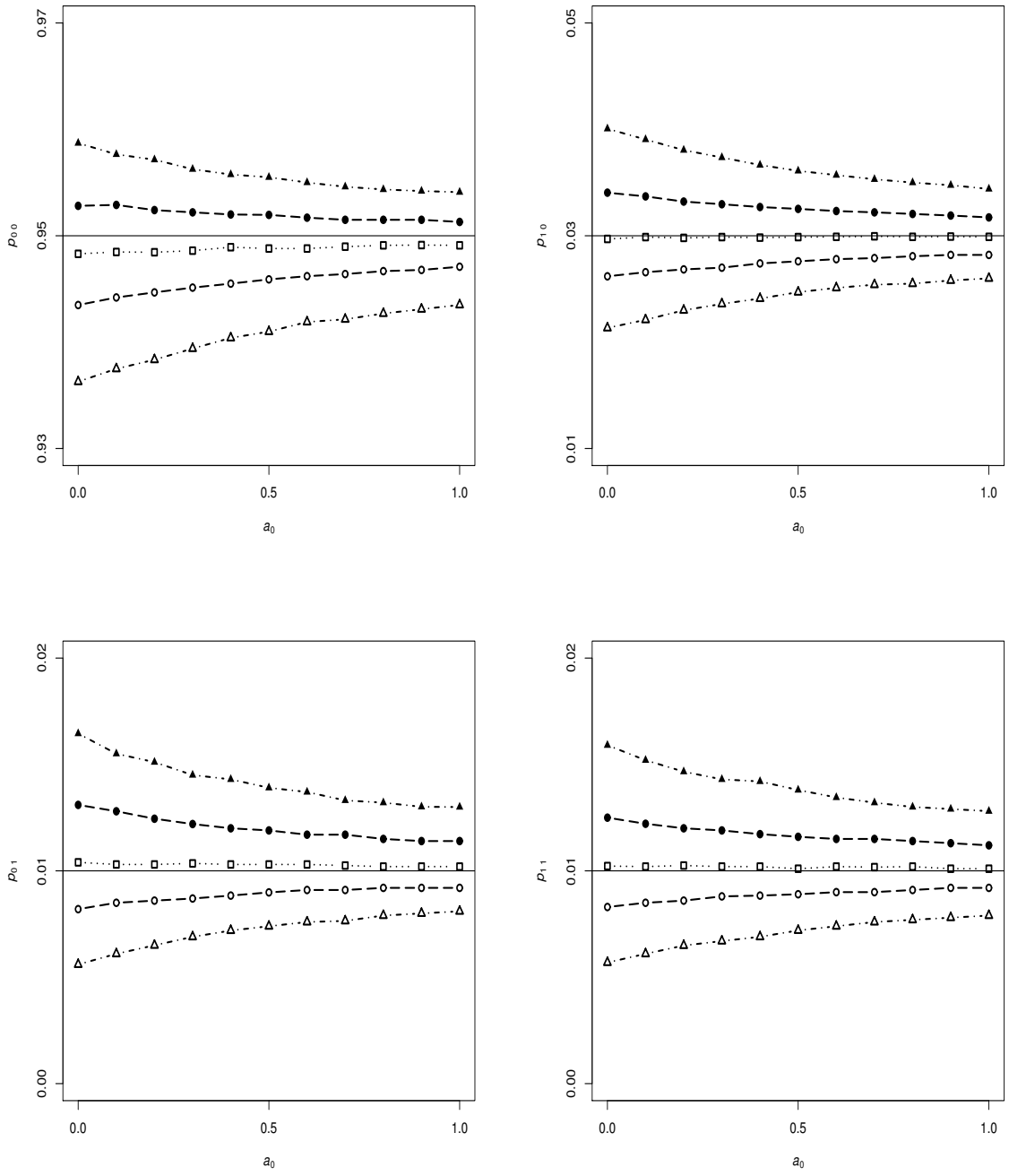

Figure A.7: Simulation results with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $\mathbf{p}$ are provided. Informative priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(109.0, 6.7)$ and $S_{p:j} \sim \text{beta}(55.2, 1.6)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

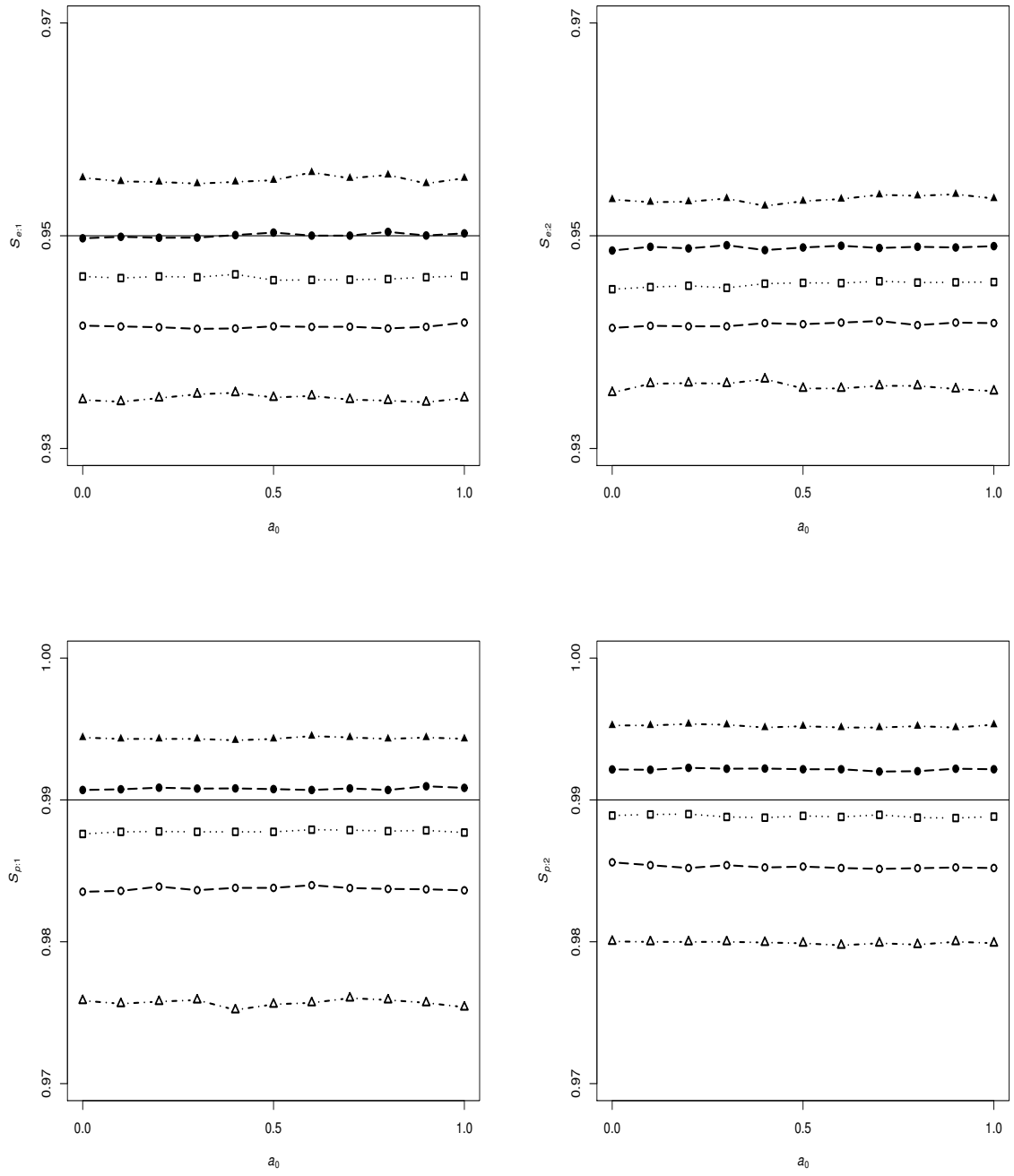

Figure A.8: Simulation results with $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$, $N = 1000$ individuals, $S_{e:j} = 0.95$, and $S_{p:j} = 0.99$, for $j = 1, 2$. The 5th (bottom), 25th, 50th (median), 75th, and 95th (top) percentiles of the $B = 500$ posterior median estimates of $S_{e:j}$ and $S_{p:j}$ are provided. Informative priors for $S_{e:j}$ and $S_{p:j}$ are used; i.e., $S_{e:j} \sim \text{beta}(109.0, 6.7)$ and $S_{p:j} \sim \text{beta}(55.2, 1.6)$. The precision parameter a_0 increases from 0 (no historical information about $\mathbf{p}$ provided) to 1 by increments of 0.1.

A.3 Comparison of Bayesian and ML estimates under misspecified assay accuracies.

At the request of an anonymous referee, we compared our Bayesian estimates of $\mathbf{p}$ with the maximum likelihood estimates of $\mathbf{p}$ from Tebbs et al. (2013) when

- incorrect beta priors are specified for $S_{e:j}$ and $S_{p:j}$
- incorrect values of $S_{e:j}$ and $S_{p:j}$ are used to calculate the ML estimate.

We used simulation to do this comparison. We took the **true values** of $S_{e:j}$ and $S_{p:j}$ to be **0.95** and **0.99**, respectively, for $j = 1, 2$, as in the manuscript. Other simulation settings are identical to those described in Section 2.4:

- $N = 1000$ individuals; $B = 500$ Monte Carlo data sets
- $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$ and $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$
- Use of pool size c_k^* that minimizes the expected number of tests.

For the Bayesian estimates, we took $G = 3000$ Gibbs iterates after discarding 1500; we then thinned the 3000 iterates by taking every 6th one. This left us with 500 posterior draws for each Bayesian estimate. As in Section 2.5, thinning was used here so that we could make a fair comparison between the Bayesian standard errors and those for the ML estimates. To examine the impact of misspecification, we considered the following prior distributions:

1. **Mild** misspecification: $S_{e:j} \sim \text{beta}(13.5, 1)$, $S_{p:j} \sim \text{beta}(13.5, 1)$. These distributions have a median value of $S_{e:j} = S_{p:j} \approx 0.95$.
2. **Moderate** misspecification: $S_{e:j} \sim \text{beta}(6.6, 1)$, $S_{p:j} \sim \text{beta}(6.6, 1)$. These distributions have a median value of $S_{e:j} = S_{p:j} \approx 0.90$.
3. **Severe** misspecification: $S_{e:j} \sim \text{beta}(4.3, 1)$, $S_{p:j} \sim \text{beta}(4.3, 1)$. These distributions have a median value of $S_{e:j} = S_{p:j} \approx 0.85$.

4. **Extreme** misspecification: $S_{e:j} \sim \text{beta}(3.1, 1)$, $S_{p:j} \sim \text{beta}(3.1, 1)$. These distributions have a median value of $S_{e:j} = S_{p:j} \approx 0.80$.

We then compared our Bayesian estimates of $\mathbf{p}$ under these wrong priors with the ML estimates calculated at the wrong (median) values of $S_{e:j}$ and $S_{p:j}$. Table A.1 shows the results for $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$. Table A.2 shows the results for $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$. Comments are given after each table is presented.

Note that our wrong priors above were chosen to be highly variable; this might be best when sound prior information is not available. We also tried more highly concentrated wrong priors (results not shown) and observed similar findings when the simulation sample size N was larger.

Table A.1: Simulation results under prior misspecification. The true value of $\mathbf{p}$ is $\mathbf{p} = (0.80, 0.10, 0.09, 0.01)$. The **true values** of $S_{e:j}$ and $S_{p:j}$ are **0.95** and **0.99**, respectively. All quantities below are as defined in Sections 2.4-2.5. The use of "$*$" with $S^*_{e:j}$ and $S^*_{p:j}$ stresses that these are the wrong values.

	Maximum likelihood		Bayesian			
	Estimate	SE	Estimate	BSE	Estimate	BSE
Mild	$\widehat{p}_{00} = 0.817$	0.0132	$\widehat{p}_{00} = 0.802$	0.0157	$\widehat{S}_{e:1} = 0.958$	0.0283
beta$(13.5, 1)$	$\widehat{p}_{10} = 0.092$	0.0099	$\widehat{p}_{10} = 0.098$	0.0116	$\widehat{S}_{e:2} = 0.956$	0.0296
$S^*_{e:j} = S^*_{p:j} = 0.95$	$\widehat{p}_{01} = 0.083$	0.0094	$\widehat{p}_{01} = 0.089$	0.0111	$\widehat{S}_{p:1} = 0.987$	0.0079
	$\widehat{p}_{11} = 0.008$	0.0031	$\widehat{p}_{11} = 0.010$	0.0036	$\widehat{S}_{p:2} = 0.988$	0.0075
Moderate	$\widehat{p}_{00} = 0.823$	0.0143	$\widehat{p}_{00} = 0.799$	0.0159	$\widehat{S}_{e:1} = 0.952$	0.0303
beta$(6.6, 1)$	$\widehat{p}_{10} = 0.090$	0.0107	$\widehat{p}_{10} = 0.099$	0.0118	$\widehat{S}_{e:2} = 0.949$	0.0320
$S^*_{e:j} = S^*_{p:j} = 0.90$	$\widehat{p}_{01} = 0.081$	0.0102	$\widehat{p}_{01} = 0.090$	0.0113	$\widehat{S}_{p:1} = 0.988$	0.0080
	$\widehat{p}_{11} = 0.006$	0.0032	$\widehat{p}_{11} = 0.010$	0.0036	$\widehat{S}_{p:2} = 0.988$	0.0075
Severe	$\widehat{p}_{00} = 0.839$	0.0160	$\widehat{p}_{00} = 0.799$	0.0161	$\widehat{S}_{e:1} = 0.949$	0.0310
beta$(4.3, 1)$	$\widehat{p}_{10} = 0.084$	0.0120	$\widehat{p}_{10} = 0.099$	0.0119	$\widehat{S}_{e:2} = 0.947$	0.0327
$S^*_{e:j} = S^*_{p:j} = 0.85$	$\widehat{p}_{01} = 0.074$	0.0114	$\widehat{p}_{01} = 0.091$	0.0114	$\widehat{S}_{p:1} = 0.988$	0.0080
	$\widehat{p}_{11} = 0.003$	0.0034	$\widehat{p}_{11} = 0.010$	0.0036	$\widehat{S}_{p:2} = 0.989$	0.0075
Extreme	$\widehat{p}_{00} = 0.869$	0.0188	$\widehat{p}_{00} = 0.798$	0.0161	$\widehat{S}_{e:1} = 0.949$	0.0313
beta$(3.1, 1)$	$\widehat{p}_{10} = 0.070$	0.0140	$\widehat{p}_{10} = 0.099$	0.0119	$\widehat{S}_{e:2} = 0.945$	0.0332
$S^*_{e:j} = S^*_{p:j} = 0.80$	$\widehat{p}_{01} = 0.059$	0.0134	$\widehat{p}_{01} = 0.091$	0.0114	$\widehat{S}_{p:1} = 0.988$	0.0079
	$\widehat{p}_{11} = 0.001$	0.0030	$\widehat{p}_{11} = 0.011$	0.0037	$\widehat{S}_{p:2} = 0.989$	0.0076

Remarks:

- As the level of misspecification increases, the ML estimates of $\mathbf{p}$ become more biased. However, the Bayesian point estimates of $\mathbf{p}$ are largely unaffected by prior model misspecification.

- The Bayesian point estimates of $S_{e:j}$ and $S_{p:j}$ are largely on target; i.e., our estimation procedure "recovers" the true values of $S_{e:j}$ and $S_{p:j}$ despite choosing incorrect priors.

- ML estimates have smaller variability at low levels of misspecification. At extreme levels of misspecification, Bayesian estimates can be more precise.

- Bayesian point estimates of $\mathbf{p}$ and $\boldsymbol{\delta}$ will only improve for larger N. ML estimates will become more precise when N is larger, but the bias will not disappear.

Table A.2: Simulation results under prior misspecification. The true value of $\mathbf{p}$ is $\mathbf{p} = (0.95, 0.03, 0.01, 0.01)$. The **true values** of $S_{e:j}$ and $S_{p:j}$ are **0.95** and **0.99**, respectively. All quantities below are as defined in Sections 2.4-2.5. The use of "$*$" with $S^*_{e:j}$ and $S^*_{p:j}$ stresses that these are the wrong values.

	Maximum likelihood		Bayesian			
	Estimate	SE	Estimate	BSE	Estimate	BSE
Mild	$\widehat{p}_{00} = 0.958$	0.0069	$\widehat{p}_{00} = 0.948$	0.0081	$\widehat{S}_{e:1} = 0.954$	0.0370
beta$(13.5, 1)$	$\widehat{p}_{10} = 0.026$	0.0056	$\widehat{p}_{10} = 0.029$	0.0064	$\widehat{S}_{e:2} = 0.947$	0.0459
$S^*_{e:j} = S^*_{p:j} = 0.95$	$\widehat{p}_{01} = 0.007$	0.0033	$\widehat{p}_{01} = 0.011$	0.0039	$\widehat{S}_{p:1} = 0.987$	0.0090
	$\widehat{p}_{11} = 0.009$	0.0031	$\widehat{p}_{11} = 0.011$	0.0034	$\widehat{S}_{p:2} = 0.989$	0.0069
Moderate	$\widehat{p}_{00} = 0.970$	0.0071	$\widehat{p}_{00} = 0.947$	0.0084	$\widehat{S}_{e:1} = 0.942$	0.0437
beta$(6.6, 1)$	$\widehat{p}_{10} = 0.019$	0.0059	$\widehat{p}_{10} = 0.030$	0.0067	$\widehat{S}_{e:2} = 0.926$	0.0590
$S^*_{e:j} = S^*_{p:j} = 0.90$	$\widehat{p}_{01} = 0.002$	0.0025	$\widehat{p}_{01} = 0.011$	0.0041	$\widehat{S}_{p:1} = 0.987$	0.0093
	$\widehat{p}_{11} = 0.009$	0.0032	$\widehat{p}_{11} = 0.011$	0.0035	$\widehat{S}_{p:2} = 0.989$	0.0070
Severe	$\widehat{p}_{00} = 0.982$	0.0066	$\widehat{p}_{00} = 0.946$	0.0086	$\widehat{S}_{e:1} = 0.937$	0.0467
beta$(4.3, 1)$	$\widehat{p}_{10} = 0.009$	0.0060	$\widehat{p}_{10} = 0.030$	0.0068	$\widehat{S}_{e:2} = 0.916$	0.0644
$S^*_{e:j} = S^*_{p:j} = 0.85$	$\widehat{p}_{01} = 0.001$	0.0007	$\widehat{p}_{01} = 0.011$	0.0043	$\widehat{S}_{p:1} = 0.987$	0.0094
	$\widehat{p}_{11} = 0.008$	0.0032	$\widehat{p}_{11} = 0.011$	0.0036	$\widehat{S}_{p:2} = 0.989$	0.0071
Extreme	$\widehat{p}_{00} = 0.991$	0.0053	$\widehat{p}_{00} = 0.946$	0.0087	$\widehat{S}_{e:1} = 0.933$	0.0482
beta$(3.1, 1)$	$\widehat{p}_{10} = 0.004$	0.0042	$\widehat{p}_{10} = 0.030$	0.0068	$\widehat{S}_{e:2} = 0.912$	0.0668
$S^*_{e:j} = S^*_{p:j} = 0.80$	$\widehat{p}_{01} = 0.000$	0.0004	$\widehat{p}_{01} = 0.012$	0.0044	$\widehat{S}_{p:1} = 0.987$	0.0094
	$\widehat{p}_{11} = 0.005$	0.0035	$\widehat{p}_{11} = 0.011$	0.0036	$\widehat{S}_{p:2} = 0.989$	0.0071

Remarks:

- As the level of misspecification increases, the ML estimates of $\mathbf{p}$ become more biased. However, the Bayesian point estimates of $\mathbf{p}$ are largely unaffected by prior model misspecification.

- Bayesian point estimates of $S_{e:j}$ can be slightly below the nominal level (especially for the second disease where the marginal prevalence is smaller); estimates of $S_{p:j}$ are generally on target.

- ML estimates have smaller variability in this case, but the bias associated with ML estimates can be large especially when misspecification is severe or extreme.

- Bayesian point estimates of $\mathbf{p}$ and $\boldsymbol{\delta}$ will only improve for larger N. ML estimates will become more precise when N is larger, but the bias will not disappear.

A.4 Additional information on the Nebraska analysis in Section 2.5.

Calculating the 2008 historical estimate $\mathbf{p}_0$: For each gender/specimen type stratum, our 2008 historical estimate $\mathbf{p}_0$ (see Table 2.1) is calculated from the observed individual testing outcomes in 2008, after adjusting for potential misclassification. To accomplish this, we first treat the assay sensitivity $S_{e:j}$ and assay specificity $S_{p:j}$ as fixed constants (see Table 2.3); these are the values stated in the Aptima Combo 2 Assay product literature, available at `http://www.hologic.com`.

Because the 2008 data are individual testing results, we remove the k subscript in our notation and denote by

$$\begin{aligned}\mathbf{Y}_i = (Y_{i1}, Y_{i2})' &= \text{2008 testing result for } i\text{th individual (what we have)} \\ \widetilde{\mathbf{Y}}_i = (\widetilde{Y}_{i1}, \widetilde{Y}_{i2})' &= \text{2008 true status for } i\text{th individual,}\end{aligned}$$

for $i = 1, 2, ..., N_0$, where N_0 is the number of 2008 individuals in each stratum (see Table 2.1). Under the same assumptions described in Section 2.3 and Section 2.6 in the manuscript, the joint distribution of the (2008) individual testing results $\mathbf{Y}$ and the latent data $\widetilde{\mathbf{Y}}$ is given by

$$\pi(\mathbf{Y}, \widetilde{\mathbf{Y}}|\mathbf{p}) = \prod_{i=1}^{N_0} p_{00}^{(1-\widetilde{Y}_{i1})(1-\widetilde{Y}_{i2})} p_{10}^{\widetilde{Y}_{i1}(1-\widetilde{Y}_{i2})} p_{01}^{(1-\widetilde{Y}_{i1})\widetilde{Y}_{i2}} p_{11}^{\widetilde{Y}_{i1}\widetilde{Y}_{i2}} \times \left[\prod_{j=1}^{2} \prod_{i=1}^{N_0} S_{e:j}^{Y_{ij}\widetilde{Y}_{ij}} \overline{S}_{e:j}^{(1-Y_{ij})\widetilde{Y}_{ij}} S_{p:j}^{(1-Y_{ij})(1-\widetilde{Y}_{ij})} \overline{S}_{p:j}^{Y_{ij}(1-\widetilde{Y}_{ij})} \right]. \quad \text{(A.2)}$$

We assume a noninformative prior for $\mathbf{p}$ in Equation (A.2); i.e., $\mathbf{p} \sim \text{Dirichlet}(\mathbf{1}_4)$. The full conditional distribution of $\mathbf{p}$; that is, $\mathbf{p}|\widetilde{\mathbf{Y}} \sim \text{Dirichlet}(\mathbf{\Psi})$, where $\mathbf{\Psi} = (1 + \sum_{i=1}^{N_0} \widetilde{V}_{i(00)}, 1 + \sum_{i=1}^{N_0} \widetilde{V}_{i(10)}, 1 + \sum_{i=1}^{N_0} \widetilde{V}_{i(01)}, 1 + \sum_{i=1}^{N_0} \widetilde{V}_{i(11)})'$ and $\widetilde{V}_{i(uv)} = \widetilde{Y}_{i1}^{u}(1 - \widetilde{Y}_{i1})^{1-u}\widetilde{Y}_{i2}^{v}(1 - \widetilde{Y}_{i2})^{1-v}$, for $u, v \in \{0, 1\}$. The full conditional distribution of $\widetilde{\mathbf{V}}_i = (\widetilde{V}_{i(00)}, \widetilde{V}_{i(10)}, \widetilde{V}_{i(01)}, \widetilde{V}_{i(11)})'$ given $\{\mathbf{p}, \mathbf{Y}\}$ is multinomial with cell probabilities ζ_{00}^i/ζ^i, ζ_{10}^i/ζ^i, ζ_{01}^i/ζ^i, and ζ_{11}^i/ζ^i, where $\zeta_{00}^i = p_{00} \prod_{j=1}^{2} S_{p:j}^{1-Y_{ij}} \overline{S}_{p:j}^{Y_{ij}}$, $\zeta_{10}^i = p_{10} S_{e:1}^{Y_{i1}} \overline{S}_{e:1}^{1-Y_{i1}} S_{p:2}^{1-Y_{i2}} \overline{S}_{p:2}^{Y_{i2}}$, $\zeta_{01}^i = p_{01} S_{e:2}^{Y_{i2}} \overline{S}_{e:2}^{1-Y_{i2}} S_{p:1}^{1-Y_{i1}} \overline{S}_{p:1}^{Y_{i1}}$, and $\zeta_{11}^i = p_{11} \prod_{j=1}^{2} S_{e:j}^{Y_{ij}} \overline{S}_{e:j}^{1-Y_{ij}}$, where $\zeta^i = \sum_{u=0}^{1} \sum_{v=0}^{1} \zeta_{uv}^i$.

Note that by sampling $\widetilde{\mathbf{V}}_i$, we determine $\widetilde{\mathbf{Y}}_i = (\widetilde{Y}_{i1}, \widetilde{Y}_{i2})'$ uniquely as shown in Section 2.3. Using known values of $S_{e:j}$ and $S_{p:j}$, one can now sample from the conditional distributions of $\mathbf{p}$ and $\widetilde{\mathbf{V}}_i$, for $i = 1, 2, ..., N_0$, until convergence. The historical estimate $\mathbf{p}_0$ shown in Table 2.1 (separately for each gender/specimen type stratum) is the median of $G = 10000$ Gibbs iterates after discarding the first 500.

Prior selection for $S_{e:j}$ and $S_{p:j}$: We use sensitivity and specificity information published in an assay's product literature to elicit prior distributions for $S_{e:j}$ and $S_{p:j}$. This information is typically collected in pilot studies using known positive and known negative specimens. Define

$$\begin{aligned} \text{TP} &= \text{number of true positive test results} \\ \text{FN} &= \text{number of false negative test results} \\ \text{TN} &= \text{number of true negative test results} \\ \text{FP} &= \text{number of false positive test results.} \end{aligned}$$

For the Aptima Combo 2 Assay, the following information was published in its product literature, which is available at `http://www.hologic.com`. This table combines information from Table 5a (CT) and Table 9a (NG) in the product literature document.

	Stratum	TP	FN	TN	FP
Chlamydia	Male/Urine	276	6	801	12
	Male/Swab	260	11	774	20
	Female/Urine	197	11	1170	13
	Female/Swab	195	12	1154	28
Gonorrhea	Male/Urine	324	5	802	3
	Male/Swab	319	3	764	17
	Female/Urine	116	11	1347	10
	Female/Swab	126	1	1335	17

For a given infection, the following prior distributions are used:

$$\begin{aligned} S_{e:j} &\sim \text{beta}(\text{TP}+1, \text{FN}+1) \\ S_{p:j} &\sim \text{beta}(\text{TN}+1, \text{FP}+1). \end{aligned}$$

These distributions can be regarded as the posterior distributions for $S_{e:j}$ and $S_{p:j}$ had they been modeled with uniform priors before the pilot study was conducted.

To illustrate, consider the male/urine stratum. For chlamydia, the prior distributions are $S_{e:1} \sim \text{beta}(277, 7)$ and $S_{p:1} \sim \text{beta}(802, 13)$. For gonorrhea, the prior distributions are $S_{e:2} \sim \text{beta}(325, 6)$ and $S_{p:2} \sim \text{beta}(803, 4)$. Prior distributions for the other strata are found similarly and are reported in Table 2.1. During 2008-2009, the Nebraska Public Health Laboratory did not use the Aptima Combo 2 Assay; however, they currently do use it. Regardless of the specific assay used, our R code at `www.chrisbilder.com/grouptesting/WTMB` determines beta prior distributions based on pilot data like those described above.

Simulating true responses for the 2009 analysis: With the observed individual testing outcomes $\mathbf{Y}_i = (Y_{i1}, Y_{i2})'$ from 2009, the 2008 historical estimate $\mathbf{p}_0$, and known values of $S_{e:j}$ and $S_{p:j}$, we sample the 2009 true statuses $\widetilde{\mathbf{Y}}_i = (\widetilde{Y}_{i1}, \widetilde{Y}_{i2})'$ from $\text{pr}(\widetilde{Y}_{i1} = \widetilde{y}_1, \widetilde{Y}_{i2} = \widetilde{y}_2 | Y_{i1} = y_1, Y_{2k} = y_2)$, for $\widetilde{y}_1, \widetilde{y}_2, y_1, y_2 \in \{0, 1\}$. This is done as follows:

1. For each $i = 1, 2, ..., N$, we sample $\widetilde{\mathbf{V}}_i = (\widetilde{V}_{i(00)}, \widetilde{V}_{i(10)}, \widetilde{V}_{i(01)}, \widetilde{V}_{i(11)})'$ given $\{\mathbf{p}_0, \mathbf{Y}\}$ from a multinomial distribution with cell probabilities ζ^i_{00}/ζ^i, ζ^i_{10}/ζ^i, ζ^i_{01}/ζ^i, and ζ^i_{11}/ζ^i, where $\zeta^i_{00} = p_{00(0)} \prod_{j=1}^{2} S_{p:j}^{1-Y_{ij}} \overline{S}_{p:j}^{Y_{ij}}$, $\zeta^i_{10} = p_{10(0)} S_{e:1}^{Y_{i1}} \overline{S}_{e:1}^{1-Y_{i1}} S_{p:2}^{1-Y_{i2}} \overline{S}_{p:2}^{Y_{i2}}$, $\zeta^i_{01} = p_{01(0)} S_{e:2}^{Y_{i2}} \overline{S}_{e:2}^{1-Y_{i2}} S_{p:1}^{1-Y_{i1}} \overline{S}_{p:1}^{Y_{i1}}$, and $\zeta^i_{11} = p_{11(0)} \prod_{j=1}^{2} S_{e:j}^{Y_{ij}} \overline{S}_{e:j}^{1-Y_{ij}}$, where $\zeta^i = \sum_{u=0}^{1} \sum_{v=0}^{1} \zeta^i_{uv}$. Recall that $\widetilde{V}_{i(uv)} = \widetilde{Y}_{i1}^u (1 - \widetilde{Y}_{i1})^{1-u} \widetilde{Y}_{i2}^v (1 - \widetilde{Y}_{i2})^{1-v}$, for $u, v \in \{0, 1\}$.

2. For each $i = 1, 2, ..., N$, we determine $\widetilde{\mathbf{Y}}_i = (\widetilde{V}_{i(10)} + \widetilde{V}_{i(11)}, \widetilde{V}_{i(01)} + \widetilde{V}_{i(11)})' = (\widetilde{Y}_{i1}, \widetilde{Y}_{i2})'$.

3. Steps 1-2 are repeated $B = 500$ times to create 500 sets of true statuses (within each gender/specimen type stratum).

Nebraska data analysis using flat priors: Tables A.3 and A.4 summarize the Nebraska data analysis performed in Section 2.5 using $S_{e:j} \sim \text{beta}(1,1)$ and $S_{p:j} \sim \text{beta}(1,1)$.

Table A.3: Nebraska CT/NG prevalence estimation results for 2009. Bayesian estimates (Bayes) are posterior medians averaged over $B = 500$ data sets; BSE is the average of the standard deviations calculated from posterior samples of the $B = 500$ data sets. Values of $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ are used to incorporate different amounts of historical information for $\mathbf{p}$ as described in Section 2.5. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used, where $j = 1$ for CT and $j = 2$ for NG; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$. Maximum likelihood estimates, calculated from Tebbs et al. (2013), are averaged over the same 500 data sets; the entries under SE are the averaged standard errors. Stratum sample sizes N are given.

			Maximum likelihood		Bayes ($a_0 = 0$)		Bayes ($a_0 = 0.5$)		Bayes ($a_0 = 1$)	
Stratum	CT	NG	Estimate	SE	Estimate	BSE	Estimate	BSE	Estimate	BSE
	$-$	$-$	$\widehat{p}_{00} = 0.924$	0.0035	$\widehat{p}_{00} = 0.922$	0.0044	$\widehat{p}_{00} = 0.925$	0.0032	$\widehat{p}_{00} = 0.927$	0.0026
Male/Urine	$+$	$-$	$\widehat{p}_{10} = 0.061$	0.0032	$\widehat{p}_{10} = 0.062$	0.0041	$\widehat{p}_{10} = 0.061$	0.0030	$\widehat{p}_{10} = 0.061$	0.0024
$N = 6139$	$-$	$+$	$\widehat{p}_{01} = 0.008$	0.0012	$\widehat{p}_{01} = 0.008$	0.0013	$\widehat{p}_{01} = 0.007$	0.0009	$\widehat{p}_{01} = 0.007$	0.0008
	$+$	$+$	$\widehat{p}_{11} = 0.007$	0.0011	$\widehat{p}_{11} = 0.007$	0.0011	$\widehat{p}_{11} = 0.006$	0.0008	$\widehat{p}_{11} = 0.006$	0.0007
	$-$	$-$	$\widehat{p}_{00} = 0.831$	0.0091	$\widehat{p}_{00} = 0.825$	0.0131	$\widehat{p}_{00} = 0.838$	0.0084	$\widehat{p}_{00} = 0.841$	0.0067
Male/Swab	$+$	$-$	$\widehat{p}_{10} = 0.119$	0.0079	$\widehat{p}_{10} = 0.118$	0.0115	$\widehat{p}_{10} = 0.110$	0.0072	$\widehat{p}_{10} = 0.108$	0.0058
$N = 1910$	$-$	$+$	$\widehat{p}_{01} = 0.034$	0.0043	$\widehat{p}_{01} = 0.039$	0.0063	$\widehat{p}_{01} = 0.036$	0.0041	$\widehat{p}_{01} = 0.034$	0.0033
	$+$	$+$	$\widehat{p}_{11} = 0.015$	0.0029	$\widehat{p}_{11} = 0.017$	0.0035	$\widehat{p}_{11} = 0.016$	0.0026	$\widehat{p}_{11} = 0.016$	0.0022
	$-$	$-$	$\widehat{p}_{00} = 0.920$	0.0041	$\widehat{p}_{00} = 0.919$	0.0054	$\widehat{p}_{00} = 0.914$	0.0040	$\widehat{p}_{00} = 0.912$	0.0033
Female/Urine	$+$	$-$	$\widehat{p}_{10} = 0.066$	0.0038	$\widehat{p}_{10} = 0.067$	0.0051	$\widehat{p}_{10} = 0.070$	0.0037	$\widehat{p}_{10} = 0.071$	0.0030
$N = 4972$	$-$	$+$	$\widehat{p}_{01} = 0.004$	0.0010	$\widehat{p}_{01} = 0.005$	0.0013	$\widehat{p}_{01} = 0.005$	0.0010	$\widehat{p}_{01} = 0.005$	0.0008
	$+$	$+$	$\widehat{p}_{11} = 0.009$	0.0014	$\widehat{p}_{11} = 0.009$	0.0015	$\widehat{p}_{11} = 0.011$	0.0013	$\widehat{p}_{11} = 0.011$	0.0011
	$-$	$-$	$\widehat{p}_{00} = 0.949$	0.0019	$\widehat{p}_{00} = 0.949$	0.0028	$\widehat{p}_{00} = 0.949$	0.0019	$\widehat{p}_{00} = 0.948$	0.0015
Female/Swab	$+$	$-$	$\widehat{p}_{10} = 0.045$	0.0019	$\widehat{p}_{10} = 0.045$	0.0027	$\widehat{p}_{10} = 0.046$	0.0019	$\widehat{p}_{10} = 0.047$	0.0015
$N = 14530$	$-$	$+$	$\widehat{p}_{01} = 0.001$	0.0001	$\widehat{p}_{01} = 0.001$	0.0002	$\widehat{p}_{01} = 0.001$	0.0001	$\widehat{p}_{01} = 0.001$	0.0001
	$+$	$+$	$\widehat{p}_{11} = 0.005$	0.0006	$\widehat{p}_{11} = 0.005$	0.0007	$\widehat{p}_{11} = 0.005$	0.0005	$\widehat{p}_{11} = 0.005$	0.0004

Table A.4: Bayesian assay accuracy estimates from 2009. Bayesian estimates (Bayes) are posterior medians averaged over $B = 500$ data sets; BSE is the average of the standard deviations calculated from posterior samples of the $B = 500$ data sets. Values of $a_0 = 0$, $a_0 = 0.5$, and $a_0 = 1$ are used to incorporate different amounts of historical information for $\mathbf{p}$ as described in Section 2.5. Flat priors for $S_{e:j}$ and $S_{p:j}$ are used, where $j = 1$ for CT and $j = 2$ for NG; i.e., $S_{e:j} \sim \text{beta}(1, 1)$ and $S_{p:j} \sim \text{beta}(1, 1)$.

		Bayes ($a_0 = 0$)		Bayes ($a_0 = 0.5$)		Bayes ($a_0 = 1$)	
Stratum	Accuracy	Estimate	BSE	Estimate	BSE	Estimate	BSE
	$S_{e:1} = 0.979$	$\widehat{S}_{e:1} = 0.975$	0.0159	$\widehat{S}_{e:1} = 0.977$	0.0140	$\widehat{S}_{e:1} = 0.978$	0.0135
Male/Urine	$S_{e:2} = 0.985$	$\widehat{S}_{e:2} = 0.968$	0.0233	$\widehat{S}_{e:2} = 0.978$	0.0191	$\widehat{S}_{e:2} = 0.982$	0.0172
$N = 6139$	$S_{p:1} = 0.985$	$\widehat{S}_{p:1} = 0.985$	0.0050	$\widehat{S}_{p:1} = 0.984$	0.0047	$\widehat{S}_{p:1} = 0.984$	0.0045
	$S_{p:2} = 0.996$	$\widehat{S}_{p:2} = 0.996$	0.0015	$\widehat{S}_{p:2} = 0.996$	0.0015	$\widehat{S}_{p:2} = 0.996$	0.0014
	$S_{e:1} = 0.959$	$\widehat{S}_{e:1} = 0.957$	0.0263	$\widehat{S}_{e:1} = 0.971$	0.0202	$\widehat{S}_{e:1} = 0.974$	0.0185
Male/Swab	$S_{e:2} = 0.991$	$\widehat{S}_{e:2} = 0.935$	0.0425	$\widehat{S}_{e:2} = 0.952$	0.0341	$\widehat{S}_{e:2} = 0.958$	0.0314
$N = 1910$	$S_{p:1} = 0.975$	$\widehat{S}_{p:1} = 0.972$	0.0116	$\widehat{S}_{p:1} = 0.966$	0.0104	$\widehat{S}_{p:1} = 0.965$	0.0099
	$S_{p:2} = 0.978$	$\widehat{S}_{p:2} = 0.985$	0.0061	$\widehat{S}_{p:2} = 0.983$	0.0055	$\widehat{S}_{p:2} = 0.982$	0.0054
	$S_{e:1} = 0.947$	$\widehat{S}_{e:1} = 0.946$	0.0199	$\widehat{S}_{e:1} = 0.935$	0.0185	$\widehat{S}_{e:1} = 0.932$	0.0177
Female/Urine	$S_{e:2} = 0.913$	$\widehat{S}_{e:2} = 0.899$	0.0449	$\widehat{S}_{e:2} = 0.881$	0.0449	$\widehat{S}_{e:2} = 0.874$	0.0447
$N = 4972$	$S_{p:1} = 0.989$	$\widehat{S}_{p:1} = 0.989$	0.0057	$\widehat{S}_{p:1} = 0.992$	0.0049	$\widehat{S}_{p:1} = 0.993$	0.0046
	$S_{p:2} = 0.993$	$\widehat{S}_{p:2} = 0.993$	0.0023	$\widehat{S}_{p:2} = 0.994$	0.0022	$\widehat{S}_{p:2} = 0.994$	0.0021
	$S_{e:1} = 0.942$	$\widehat{S}_{e:1} = 0.942$	0.0163	$\widehat{S}_{e:1} = 0.937$	0.0148	$\widehat{S}_{e:1} = 0.935$	0.0143
Female/Swab	$S_{e:2} = 0.992$	$\widehat{S}_{e:2} = 0.960$	0.0278	$\widehat{S}_{e:2} = 0.968$	0.0250	$\widehat{S}_{e:2} = 0.970$	0.0238
$N = 14530$	$S_{p:1} = 0.976$	$\widehat{S}_{p:1} = 0.975$	0.0043	$\widehat{S}_{p:1} = 0.977$	0.0039	$\widehat{S}_{p:1} = 0.977$	0.0037
	$S_{p:2} = 0.987$	$\widehat{S}_{p:2} = 0.988$	0.0015	$\widehat{S}_{p:2} = 0.987$	0.0015	$\widehat{S}_{p:2} = 0.988$	0.0015

Appendix B

Chapter 3 supplementary materials

B.1 E-step and Gibbs sampler for the EM algorithm in Section 3.2.

We herein provide closed-form expressions for the expectations $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ and $E(\widetilde{Y}_{ij}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ given in Section 3.2.

Expectation in Step 1 of the EM algorithm in Section 3.2:

$$E\left(I_{jk}\middle|\mathcal{D}_D, \boldsymbol{\theta}^{(d)}\right) = \left(\frac{\mu^{(d)}_{\widetilde{z}:jk}}{\sum_{s=0}^{c_j}\mu^{(d)}_{\widetilde{z}:js}}\right)^{Z_j}\left[\frac{\left\{1-h\left(k,c_j,\lambda^{(d)}\right)\right\}\text{pr}\left(\sum_{i=1}^{c_j}\widetilde{Y}_{ij}=k\right)^{(d)}}{1-p_j^{(d)}}\right]^{1-Z_j}$$

where

$$\begin{aligned}
p_{ij}^{(d)} &= g^{-1}\left(\mathbf{x}_{ij}'\boldsymbol{\beta}^{(d)}\right)\\
p_j^{(d)} &= (1-S_p)\prod_{i=1}^{c_j}\left(1-p_{ij}^{(d)}\right)+\sum_{k=1}^{c_j}h\left(k,c_j,\lambda^{(d)}\right)\text{pr}\left(\sum_{i=1}^{c_j}\widetilde{Y}_{ij}=k\right)^{(d)}\\
\mu^{(d)}_{\widetilde{z}:js} &= \sum_{\widetilde{y}_{2j}=0}^{1}\cdots\sum_{\widetilde{y}_{c_jj}=0}^{1}\Bigg[I\left(s-1\le\sum_{i=2}^{c_j}\widetilde{y}_{ij}\le s\right)(1-S_p)^{I\left(\sum_{i=1}^{c_j}\widetilde{y}_{ij}=0\right)}\\
&\quad\times\left\{\prod_{u=1}^{c_j}h\left(u,c_j,\lambda^{(d)}\right)^{I\left(\sum_{i=1}^{c_j}\widetilde{y}_{ij}=u\right)}\right\}\left\{\prod_{i=1}^{c_j}\left(p_{ij}^{(d)}\right)^{\widetilde{y}_{ij}}\left(1-p_{ij}^{(d)}\right)^{1-\widetilde{y}_{ij}}\right\}\\
&\quad\times\prod_{i=1}^{c_j}\left\{S_e^{\widetilde{y}_{ij}}(1-S_p)^{1-\widetilde{y}_{ij}}\right\}^{Y_{ij}}\left\{(1-S_e)^{\widetilde{y}_{ij}}S_p^{1-\widetilde{y}_{ij}}\right\}^{1-Y_{ij}}\Bigg],
\end{aligned}$$

$\widetilde{y}_{1j}=s-\sum_{i=2}^{c_j}\widetilde{y}_{ij}$, $\widetilde{y}_{ij}\in\{0,1\}$, and $\text{pr}(\sum_{i=1}^{c_j}\widetilde{Y}_{ij}=k)^{(d)}$ is $\text{pr}(\sum_{i=1}^{c_j}\widetilde{Y}_{ij}=k)$ evaluated at $\boldsymbol{\beta}=\boldsymbol{\beta}^{(d)}$. Note that the expression of $\mu^{(d)}_{\widetilde{z}:js}$ involves 2^{c_j-1} terms. Therefore, evaluating $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ in closed form for large pool sizes can be computationally intractable. To obviate this difficulty, we approximate $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ using the Gibbs sampler presented below.

Expectation in Step 3 of the EM algorithm in Section 3.2:

$$E\left(\widetilde{Y}_{ij}|\mathcal{D}_D,\boldsymbol{\theta}^{(d)}\right) = \left(\frac{\mu^{(d)}_{1\widetilde{y}:ij}}{\mu^{(d)}_{0\widetilde{y}:ij}+\mu^{(d)}_{1\widetilde{y}:ij}}\right)^{Z_j} \times\left[\frac{p^{(d)}_{ij}\sum_{k=0}^{c_j-1}\left\{1-h(k+1,c_j,\lambda^{(d)})\right\}\text{pr}\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=k\right)^{(d)}}{1-p^{(d)}_j}\right]^{1-Z_j},$$

where

$$\mu^{(d)}_{0\widetilde{y}:ij} = S_p^{1-Y_{ij}}\left(1-S_p\right)^{Y_{ij}}\left(1-p^{(d)}_{ij}\right) \times\left\{(1-S_p)\prod_{i'\neq i}\left(1-p^{(d)}_{ij}\right)+\sum_{k=1}^{c_j-1}h\left(k,c_j,\lambda^{(d)}\right)\text{pr}\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=k\right)^{(d)}\right\} \quad \text{(B.1)}$$

$$\mu^{(d)}_{1\widetilde{y}:ij} = S_e^{Y_{ij}}\left(1-S_e\right)^{1-Y_{ij}}p^{(d)}_{ij}\sum_{k=0}^{c_j-1}h\left(k+1,c_j,\lambda^{(d)}\right)\text{pr}\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=k\right)^{(d)}. \quad \text{(B.2)}$$

Gibbs sampler to approximate $E(I_{jk}|\mathcal{D}_D,\boldsymbol{\theta}^{(d)})$:

Let $\widetilde{\mathbf{Y}}_{(-i)j}=(\widetilde{Y}_{1j},...,\widetilde{Y}_{i-1,j},\widetilde{Y}_{i+1,j},...,\widetilde{Y}_{c_jj})'$ denote the collection of true statuses of the individuals in the jth pool except the ith one. Also, let $\mathbf{Y}_j=(Y_{1j},Y_{2j},...,Y_{c_jj})'$ denote the individual retest results from the jth pool. At the current estimate $\boldsymbol{\theta}^{(d)}$, $\widetilde{Y}_{ij}|\{\mathbf{Z},\mathbf{Y}_j,\widetilde{\mathbf{Y}}_{(-i)j}\}$ follows a Bernoulli distribution with success and failure probabilities $\zeta_1^{ij(d)}/\zeta_+^{ij(d)}$ and $\zeta_0^{ij(d)}/\zeta_+^{ij(d)}$, respectively, where $\zeta_+^{ij(d)}=\zeta_1^{ij(d)}+\zeta_0^{ij(d)}$,

$$\zeta_1^{ij(d)} = p^{(d)}_{ij}\left\{S_e^{Y_{ij}}(1-S_e)^{1-Y_{ij}}\prod_{k=1}^{c_j}\gamma^{(d)}_{ijk}\right\}^{Z_j}\left\{\prod_{k=1}^{c_j}\overline{\gamma}^{(d)}_{ijk}\right\}^{1-Z_j},$$

$$\zeta_0^{ij(d)} = \left(1-p^{(d)}_{ij}\right)\left\{(1-S_p)^{Y_{ij}+I\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=0\right)}S_p^{1-Y_{ij}}\prod_{k=1}^{c_j}\gamma^{(d)}_{ijk}\right\}^{Z_j} \times\left\{S_p^{I\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=0\right)}\prod_{k=1}^{c_j}\overline{\gamma}^{(d)}_{ijk}\right\}^{1-Z_j},$$

$\gamma^{(d)}_{ijk} = h(k,c_j,\lambda^{(d)})^{I\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=k-1\right)}$, and $\overline{\gamma}^{(d)}_{ijk} = \left\{1-h(k,c_j,\lambda^{(d)})\right\}^{I\left(\sum_{i'\neq i}\widetilde{Y}_{i'j}=k-1\right)}$. We sample $\widetilde{Y}_{ij}|\{\mathbf{Z},\mathbf{Y}_j,\widetilde{\mathbf{Y}}_{(-i)j}\}$ from a Bernoulli distribution for each $i=1,2,...,c_j$ and $j=1,2,...,J$ and repeat this procedure a large number of times to approximate $E(I_{jk}|\mathcal{D}_D,\boldsymbol{\theta}^{(d)})$. The Gibbs sampler works well overall. We did not observe

any major differences between the exact and approximate approaches to calculate $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$ when it was feasible to do this comparison. Now, we present the complete EM algorithm which implements the Gibbs sampler for approximating $E(I_{jk}|\mathcal{D}_D, \boldsymbol{\theta}^{(d)})$. This EM algorithm works for any large pool size c_j.

EM ALGORITHM

1. Specify $\boldsymbol{\theta}^{(0)}$, the number of Gibbs iterates G, and the burn-in period a. Set $d = 0$.

2. Steps for estimating λ:

   a) Initialize $\widetilde{Y}_{ij}^{(0)} = 0$ for $i = 1, 2, ..., c_j$ and $j = 1, 2, ..., J$. Aggregate the $\widetilde{Y}_{ij}^{(0)}$'s into $\widetilde{\mathbf{Y}}^{(0)}$. Set $s = 0$.

   b) For each $i = 1, 2, ..., c_j$ and $j = 1, 2, ..., J$, sample

   $$\widetilde{Y}_{ij}^{(s+1)} | \left\{ \mathbf{Z}, \mathbf{Y}_j, \widetilde{\mathbf{Y}}_{(-i)j}^{(s)} \right\} \sim \text{Bernoulli} \left(\zeta_1^{ij(d)} / \zeta_+^{ij(d)} \right),$$

   where $\widetilde{\mathbf{Y}}_{(-i)j}^{(s)} = \left(\widetilde{Y}_{1j}^{(s)}, ..., \widetilde{Y}_{i-1,j}^{(s)}, \widetilde{Y}_{i+1,j}^{(s)}, ..., \widetilde{Y}_{c_j j}^{(s)} \right)'$. Aggregate the $\widetilde{Y}_{ij}^{(s+1)}$'s into $\widetilde{\mathbf{Y}}^{(s+1)}$.

   c) Set $s = s + 1$ and repeat (b) while $s < G$.

   d) For $j = 1, 2, ..., J$ and $k = 1, 2, ..., c_j$ calculate

   $$\overline{I}_{jk} = \frac{1}{G - a} \sum_{s=a+1}^{G} I \left(\sum_{i=1}^{c_j} \widetilde{Y}_{ij}^{(s)} = k \right).$$

   We use $\overline{I}_{jk}$ as an approximation of $E\left(I_{jk} \middle| \mathcal{D}_D, \boldsymbol{\theta}^{(d)} \right)$.

   e) Find $\lambda^{(d+1)} = \arg\max_\lambda \mathcal{T}_2(\lambda, \mathcal{D}_D, \boldsymbol{\theta}^{(d)})$.

3. Steps for estimating $\boldsymbol{\beta}$:

   a) Evaluate $E\left\{ \widetilde{Y}_{ij} | \mathcal{D}_D, \left(\boldsymbol{\beta}^{(d)}, \lambda^{(d+1)} \right)' \right\}$ using the closed-form expressions provided in the E-Step.

b) Find $\boldsymbol{\beta}^{(d+1)} = \arg\max_{\boldsymbol{\beta}} \mathcal{T}_1\{\boldsymbol{\beta}, \mathcal{D}_D, (\boldsymbol{\beta}^{(d)\prime}, \lambda^{(d+1)})'\}$.

4. Set $\boldsymbol{\theta}^{(d+1)} = (\boldsymbol{\beta}^{(d+1)}, \lambda^{(d+1)})'$ and $d = d+1$. Repeat the steps above until $\boldsymbol{\theta}^{(d)}$ converges.

B.2 Covariance matrix estimation using Louis's method.

Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\lambda})'$ denote the estimate of $\boldsymbol{\theta}$ at convergence. As discussed in Section 3.2, the estimated standard errors are calculated via $\mathcal{I}(\widehat{\boldsymbol{\theta}})^{-1}$, where

$$\mathcal{I}(\boldsymbol{\theta}) = -\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \operatorname{cov}\left\{ \frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \boldsymbol{\theta}} \middle| \mathcal{D}_D, \boldsymbol{\theta} \right\}.$$

We now present an explicit expression for $\mathcal{I}(\boldsymbol{\theta})$ when $g(t) = \log\{t/(1-t)\}$ and h is the submodel provided in (3.8). For notational convenience, define $U_{jk} = E(I_{jk} | \mathcal{D}_D, \boldsymbol{\theta})$ and $V_{ij} = E(\widetilde{Y}_{ij} | \mathcal{D}_D, \boldsymbol{\theta})$. The components of $\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'$ are given by

$$\begin{aligned}
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\sum_{j=1}^{J} \sum_{i=1}^{c_j} p_{ij}(1 - p_{ij}) \mathbf{x}_{ij} \mathbf{x}_{ij}' \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \lambda} &= 0 \\
\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta})}{\partial \lambda^2} &= -\sum_{j=1}^{J} \sum_{k=1}^{c_j} U_{jk}\, h(k, c_j, \lambda) \left\{1 - h(k, c_j, \lambda)\right\} \tau(k, c_j)^2,
\end{aligned}$$

where $\tau(k, c_j) = (k - c_j)/c_j$. The components of $\operatorname{cov}\{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}}) / \partial \boldsymbol{\theta} | \mathcal{D}_D, \boldsymbol{\theta}\}$ are given by

$$\begin{aligned}
\operatorname{cov}\left(\frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \boldsymbol{\beta}}, \frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \boldsymbol{\beta}} \middle| \mathcal{D}_D, \boldsymbol{\theta} \right) &= \sum_{j=1}^{J} \sum_{i=1}^{c_j} V_{ij}\,(1 - V_{ij})\, \mathbf{x}_{ij} \mathbf{x}_{ij}' \\
\operatorname{cov}\left(\frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \boldsymbol{\beta}}, \frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \lambda} \middle| \mathcal{D}_D, \boldsymbol{\theta} \right) &= \sum_{j=1}^{J} \sum_{k=1}^{c_j} \delta_{jk} \sum_{i=1}^{c_j} \left(\mu_{1:ijk}^{Z_j}\, \mu_{0:ijk}^{1-Z_j} - U_{jk}\, V_{ij} \right) \mathbf{x}_{ij} \\
\operatorname{cov}\left(\frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \lambda}, \frac{\partial l_c(\boldsymbol{\theta} | \mathcal{D}_D, \widetilde{\mathbf{Y}})}{\partial \lambda} \middle| \mathcal{D}_D, \boldsymbol{\theta} \right) &= \sum_{j=1}^{J} \sum_{s=1}^{c_j} \sum_{t=1}^{c_j} \Big[\delta_{js} \delta_{jt} \left\{ U_{js}(1 - U_{js}) \right\}^{I(s=t)} \\
&\qquad \times \left(-U_{js}\, U_{jt} \right)^{I(s \neq t)} \Big],
\end{aligned}$$

where

$$\begin{aligned}
\delta_{jk} &= [Z_j \{1 - h(k, c_j, \lambda)\} - (1 - Z_j)\, h(k, c_j, \lambda)]\, \tau(k, c_j) \\
\mu_{1:ijk} &= \frac{h(k, c_j, \lambda)\, S_e^{Y_{ij}} (1 - S_e)^{1 - Y_{ij}}\, p_{ij}\, \mathrm{pr}\left(\sum_{i' \neq i}^{c_j} \widetilde{Y}_{i'j} = k - 1\right)}{\mu_{0\widetilde{y}:ij} + \mu_{1\widetilde{y}:ij}} \\
\mu_{0:ijk} &= \frac{\{1 - h(k, c_j, \lambda)\}\, p_{ij}\, \mathrm{pr}(\sum_{i' \neq i}^{c_j} \widetilde{Y}_{i'j} = k - 1)}{S_p \prod_{i=1}^{c_j} (1 - p_{ij}) + \sum_{k=1}^{c_j} \{1 - h(k, c_j, \lambda)\}\, \mathrm{pr}\left(\sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right)},
\end{aligned}$$

and $\mu_{0\widetilde{y}:ij}$ and $\mu_{1\widetilde{y}:ij}$ are given in Equations (B.1) and (B.2).

B.3 Observed likelihood function for Dorfman decoding.

In Section 3.3, we developed a likelihood ratio test to detect dilution. We now show how one can evaluate the observed likelihood function for Dorfman decoding. Let $\mathbf{Y}_j = (Y_{1j}, Y_{2j}, ..., Y_{c_j j})'$ and $\widetilde{\mathbf{Y}}_j = (\widetilde{Y}_{1j}, \widetilde{Y}_{2j}, ..., \widetilde{Y}_{c_j j})'$. Then, we have

$$\begin{aligned}
p_{0:j} &= \mathrm{pr}(Z_j = 0) = S_p \prod_{i=1}^{c_j} (1 - p_{ij}) + \sum_{k=1}^{c_j} \{1 - h(k, c_j, \lambda)\}\ \mathrm{pr}\left(\sum_{i=1}^{c_j} \widetilde{Y}_{ij} = k\right) \\
p_{1:j} &= \mathrm{pr}(Z_j = 1, \mathbf{Y}_j = \mathbf{y}_j) = \sum_{\widetilde{y}_{1j}=0}^{1} ... \sum_{\widetilde{y}_{c_j j}=0}^{1} f_1(\widetilde{\mathbf{y}}_j, \lambda)\, f_2(\mathbf{y}_j, \widetilde{\mathbf{y}}_j)\, f_3(\widetilde{\mathbf{y}}_j, \boldsymbol{\beta}),
\end{aligned} \tag{B.3}$$

where

$$\begin{aligned}
f_1(\widetilde{\mathbf{y}}_j, \lambda) &= (1 - S_p)^{I(\sum_{i=1}^{c_j} \widetilde{y}_{ij} = 0)} \prod_{k=1}^{c_j} h\,(k, c_j, \lambda)^{I(\sum_{i=1}^{c_j} \widetilde{y}_{ij} = k)} \\
f_2(\mathbf{y}_j, \widetilde{\mathbf{y}}_j) &= \prod_{i=1}^{c_j} \left\{ S_e^{\widetilde{y}_{ij}} (1 - S_p)^{1 - \widetilde{y}_{ij}} \right\}^{y_{ij}} \left\{ (1 - S_e)^{\widetilde{y}_{ij}} S_p^{1 - \widetilde{y}_{ij}} \right\}^{1 - y_{ij}} \\
f_3(\widetilde{\mathbf{y}}_j, \boldsymbol{\beta}) &= \prod_{i=1}^{c_j} p_{ij}^{\widetilde{y}_{ij}}\, (1 - p_{ij})^{(1 - \widetilde{y}_{ij})},
\end{aligned}$$

where $\widetilde{\mathbf{y}}_j = (\widetilde{y}_{1j}, \widetilde{y}_{2j}, ..., \widetilde{y}_{c_j j})'$. Equation (B.3) is derived under the assumption, mentioned in Section 3.2, that testing responses are independent conditional on their true statuses. The observed likelihood function is given by

$$L(\boldsymbol{\theta} | \mathcal{D}_D) = \prod_{j=1}^{J} p_{1:j}^{Z_j}\, p_{0:j}^{1 - Z_j}.$$

To perform the hypothesis test in Sections 3.4 and 3.5, we used the exact approach presented above for evaluating $L(\boldsymbol{\theta} | \mathcal{D}_D)$ at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. However, this approach can be

computationally infeasible when c_j is very large because Equation (B.3) involves 2^{c_j} terms. To avoid this computational difficulty, one can approximate $p_{1:j}$ in Equation (B.3) empirically as follows. Note that

$$\begin{aligned} p_{1:j} &= \text{pr}(Z_j = 1, \mathbf{Y}_j = \mathbf{y}_j) \\ &= \int_{\widetilde{\mathbf{y}}_j} \text{pr}(Z_j = 1, \mathbf{Y}_j = \mathbf{y}_j | \widetilde{\mathbf{Y}}_j = \widetilde{\mathbf{y}}_j) \, \text{pr}(\widetilde{\mathbf{Y}}_j = \widetilde{\mathbf{y}}_j) \, d\widetilde{\mathbf{y}}_j \\ &= E_{\widetilde{\mathbf{Y}}_j} \left\{ f_1(\widetilde{\mathbf{Y}}_j, \lambda) \, f_2(\mathbf{y}_j, \widetilde{\mathbf{Y}}_j) \right\}. \end{aligned}$$

At $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ (the MLE of $\boldsymbol{\theta}$), let $\{\widetilde{\mathbf{y}}_j^{(1)}, \widetilde{\mathbf{y}}_j^{(2)}, ..., \widetilde{\mathbf{y}}_j^{(M)}\}$ denote a Markov Chain Monte Carlo (MCMC) sample from $\text{pr}(\widetilde{\mathbf{Y}}_j = \widetilde{\mathbf{y}}_j)$. The empirical mean is given by

$$\overline{p}_{1:j} = \frac{1}{M} \sum_{m=1}^{M} f_1(\widetilde{\mathbf{y}}_j^{(m)}, \lambda) \, f_2(\mathbf{y}_j, \widetilde{\mathbf{y}}_j^{(m)}),$$

where $f_1(\widetilde{\mathbf{y}}_j, \lambda)$ and $f_2(\mathbf{y}_j, \widetilde{\mathbf{y}}_j)$ are defined above. For sufficiently large M, one can use $\overline{p}_{1:j}$ as an approximation of $p_{1:j}$ for any (large) pool size c_j. Note that one can sample $\{\widetilde{\mathbf{y}}_j^{(1)}, \widetilde{\mathbf{y}}_j^{(2)}, ..., \widetilde{\mathbf{y}}_j^{(M)}\}$ using the Gibbs sampler presented in Appendix B.1.

B.4 Additional information about dilution submodels.

Equation (3.8) presents the parametric submodel we assume in Section 3.4:

$$h(k, c_j, \lambda) = \frac{\exp\{\lambda \, \tau(k, c_j)\}}{S_e^{-1} + \exp\{\lambda \, \tau(k, c_j)\} - 1}. \tag{B.4}$$

We obtain this function by manipulating the cumulative distribution function of a logistic random variable with the assumption mentioned in Section 3.2 that $S_e = h(k, c_j, \lambda)$ for $c_j = 1$ (individual testing). The function initially had the form of $\text{logit}\{h(k, c_j, \lambda)\} = \lambda_0 + \lambda k / c_j$, where λ_0 and λ are both unknown scalar constants. Next, we set $S_e = h(k, c_j, \lambda)$ for $c_j = 1$ so that $S_e = \exp(\lambda_0 + \lambda)/\{1 + \exp(\lambda_0 + \lambda)\}$ and then solve the equation for λ_0; finally we plug-in λ_0 back to the original expression of $h(k, c_j, \lambda)$ to obtain the submodel in (B.4).

To study the robustness of our proposed method to submodel misspecification, we used the following submodels:

$$\begin{aligned}
\text{HS:} \quad h(k, c_j, \lambda) &= \frac{S_e\, k}{k + (c_j - k)\lambda} \\
\text{Probit:} \quad h(k, c_j, \lambda) &= \Phi\left\{\Phi^{-1}(S_e) + \tau_1(k, c_j)\lambda\right\} \\
\text{Cloglog:} \quad h(k, c_j, \lambda) &= 1 - \exp\left[-\exp\left[\log\{-\log(1 - S_e)\} + \tau_1(k, c_j)\lambda\right]\right],
\end{aligned}$$

where $\tau_1(k, c_j) = (k - c_j)/c_j$, $\lambda \geq 0$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. These submodels obey the properties of $h(k, c_j, \lambda)$ defined in Section 3.2. The first one, which we call HS, was originally proposed by Hung and Swallow (1999). The latter two functions are derived by manipulating the cumulative distribution functions associated with probit and complementary log-log models similarly to how we derived the submodel in Equation (B.4) from the logit model.

The robustness study is done in two steps. First, we simulate group testing data using one of the submodels HS, Probit, and Cloglog. Second, we fit our model in Equation (B.4) specified for $h(k, c_j, \lambda)$. The robustness study results presented next use the parameter configurations listed below.

- Moderate misclassification (Figures B.1-B.6): $\lambda = 0.05, 1.7, 1.2$ for HS, Probit, and Cloglog submodels in this order.

- Severe misclassification (Figures B.7-B.12): $\lambda = 0.11, 2.3, 1.7$ for HS, Probit, and Cloglog submodels in this order.

B.5 Simulation results from Section 3.4.

We present additional simulation results from Section 3.4.

- Table B.1: Estimation results for homogeneous pooling. The same results for random pooling are presented in Section 3.4.
- Table B.2: Power properties of the hypothesis test with misspecified submodels.
- Figure B.1: Robustness study for **random pooling** and **moderate misclassification**; data simulated using HS.
- Figure B.2: Robustness study for **homogeneous pooling** and **moderate misclassification**; data simulated using HS.
- Figure B.3: Robustness study for **random pooling** and **moderate misclassification**; data simulated using Probit.
- Figure B.4: Robustness study for **homogeneous pooling** and **moderate misclassification**; data simulated using Probit.
- Figure B.5: Robustness study for **random pooling** and **moderate misclassification**; data simulated using Cloglog.
- Figure B.6: Robustness study for **homogeneous pooling** and **moderate misclassification**; data simulated using Cloglog.

- Figure B.7: Robustness study for **random pooling** and **severe misclassification**; data simulated using HS.
- Figure B.8: Robustness study for **homogeneous pooling** and **severe misclassification**; data simulated using HS.
- Figure B.9: Robustness study for **random pooling** and **severe misclassification**; data simulated using Probit.
- Figure B.10: Robustness study for **homogeneous pooling** and **severe misclassification**; data simulated using Probit.
- Figure B.11: Robustness study for **random pooling** and **severe misclassification**; data simulated using Cloglog.
- Figure B.12: Robustness study for **homogeneous pooling** and **severe misclassification**; data simulated using Cloglog.

Table B.1: Simulation results for master pool testing (MPT) and Dorfman decoding (DD) with $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \lambda)' = (-3, 2, 1, \lambda)'$. "Mean" is the averaged maximum likelihood estimate and SE is the averaged standard error estimate calculated from 500 simulated data sets. Cov is the estimated coverage rate of nominal 95% Wald confidence intervals. The margin of error for the estimated coverage rate, assuming a 99% confidence level, is 0.03. Constant pool sizes c are used. Homogeneous pooling has been used for this simulation.

			Constant S_e/S_p			Dilution		
c			$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
				$\lambda = 2.6$				
5	MPT	Mean (SE)	−3.07 (0.10)	1.94 (0.14)	0.96 (0.21)	−2.81 (0.32)	2.02 (0.18)	1.00 (0.23)
		Cov	0.90	0.89	0.94	0.93	0.92	0.95
	DD	Mean (SE)	−3.08 (0.09)	2.03 (0.09)	1.03 (0.15)	−3.01 (0.10)	2.01 (0.10)	1.01 (0.15)
		Cov	0.84	0.93	0.94	0.94	0.95	0.96
10	MPT	Mean (SE)	−3.13 (0.10)	1.85 (0.18)	0.93 (0.28)	−2.82 (0.30)	2.07 (0.26)	1.05 (0.34)
		Cov	0.84	0.80	0.94	0.93	0.90	0.97
	DD	Mean (SE)	−3.11 (0.09)	2.06 (0.09)	1.02 (0.16)	−3.01 (0.10)	2.01 (0.10)	1.00 (0.16)
		Cov	0.78	0.89	0.93	0.94	0.96	0.94
				$\lambda = 3.8$				
5	MPT	Mean (SE)	−3.20 (0.10)	1.88 (0.13)	0.94 (0.21)	−2.71 (0.52)	2.00 (0.20)	1.00 (0.24)
		Cov	0.52	0.82	0.94	0.87	0.92	0.96
	DD	Mean (SE)	−3.19 (0.09)	2.07 (0.10)	1.03 (0.16)	−3.00 (0.11)	2.00 (0.10)	1.00 (0.16)
		Cov	0.47	0.89	0.94	0.95	0.96	0.97
10	MPT	Mean (SE)	−3.34 (0.11)	1.71 (0.17)	0.86 (0.28)	−2.58 (0.53)	2.03 (0.29)	1.03 (0.36)
		Cov	0.08	0.56	0.87	0.83	0.95	0.94
	DD	Mean (SE)	−3.28 (0.09)	2.15 (0.10)	1.07 (0.16)	−3.01 (0.11)	2.01 (0.11)	1.00 (0.17)
		Cov	0.14	0.71	0.91	0.96	0.96	0.96
				$\lambda = 5.0$				
5	MPT	Mean (SE)	−3.44 (0.11)	1.82 (0.14)	0.90 (0.22)	−2.60 (0.77)	1.91 (0.24)	0.95 (0.26)
		Cov	0.00	0.69	0.92	0.86	0.93	0.93
	DD	Mean (SE)	−3.43 (0.10)	2.17 (0.10)	1.08 (0.17)	−3.00 (0.12)	2.00 (0.11)	0.99 (0.17)
		Cov	0.01	0.61	0.87	0.95	0.97	0.95
10	MPT	Mean (SE)	−3.74 (0.13)	1.64 (0.17)	0.81 (0.28)	−2.65 (0.66)	1.90 (0.30)	0.95 (0.36)
		Cov	0.00	0.40	0.90	0.87	0.95	0.94
	DD	Mean (SE)	−3.64 (0.11)	2.34 (0.11)	1.17 (0.18)	−3.00 (0.13)	1.99 (0.12)	0.99 (0.18)
		Cov	0.00	0.17	0.78	0.96	0.96	0.95

Table B.2: Robustness study with misspecified submodels for master pool testing (MPT) and Dorfman decoding (DD), where $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \lambda)' = (-3, 2, 1, \lambda)'$. The proposed methods with the assumed submodel in (3.8) are fitted to the group testing data generated using the submodels HS, Probit, and Cloglog. Estimated size and power of the $\alpha = 0.05$ likelihood ratio test calculated from 500 simulated data sets. The margin of error for the estimated size when $\lambda = 0$, assuming a 99% confidence level, is 0.03. Constant pool sizes c and unequal (UE) pool sizes are used.

		HS					Probit					Cloglog				
c		$\lambda = 0$	0.02	0.04	0.06	0.08	0	0.5	1	1.5	2	0	0.4	0.8	1.2	1.6
Random pooling																
5	MPT	0.05	0.11	0.19	0.24	0.31	0.06	0.05	0.09	0.18	0.27	0.05	0.06	0.14	0.20	0.30
	DD	0.05	0.37	0.78	0.94	0.99	0.04	0.09	0.27	0.66	0.97	0.04	0.10	0.38	0.90	1.00
10	MPT	0.06	0.18	0.22	0.24	0.29	0.06	0.08	0.14	0.22	0.26	0.05	0.11	0.17	0.23	0.26
	DD	0.04	0.94	1.00	1.00	1.00	0.05	0.10	0.51	0.99	1.00	0.04	0.19	0.78	1.00	1.00
UE	MPT	0.06	0.38	0.65	0.80	0.89	0.03	0.10	0.18	0.36	0.67	0.06	0.13	0.27	0.43	0.66
	DD	0.05	0.87	1.00	1.00	1.00	0.05	0.12	0.49	0.94	1.00	0.05	0.16	0.69	1.00	1.00
Homogeneous pooling																
5	MPT	0.05	0.10	0.16	0.17	0.19	0.05	0.07	0.10	0.15	0.26	0.05	0.07	0.15	0.19	0.27
	DD	0.07	0.34	0.78	0.96	0.99	0.04	0.12	0.28	0.74	0.99	0.04	0.15	0.50	0.92	1.00
10	MPT	0.05	0.23	0.28	0.29	0.36	0.03	0.12	0.25	0.39	0.44	0.06	0.16	0.35	0.49	0.44
	DD	0.05	0.73	0.99	1.00	1.00	0.04	0.18	0.51	0.96	1.00	0.05	0.21	0.73	0.99	1.00
UE	MPT	0.06	0.20	0.21	0.24	0.25	0.04	0.14	0.25	0.32	0.37	0.05	0.15	0.30	0.41	0.39
	DD	0.04	0.77	0.99	1.00	1.00	0.06	0.13	0.54	0.95	1.00	0.04	0.23	0.75	1.00	1.00

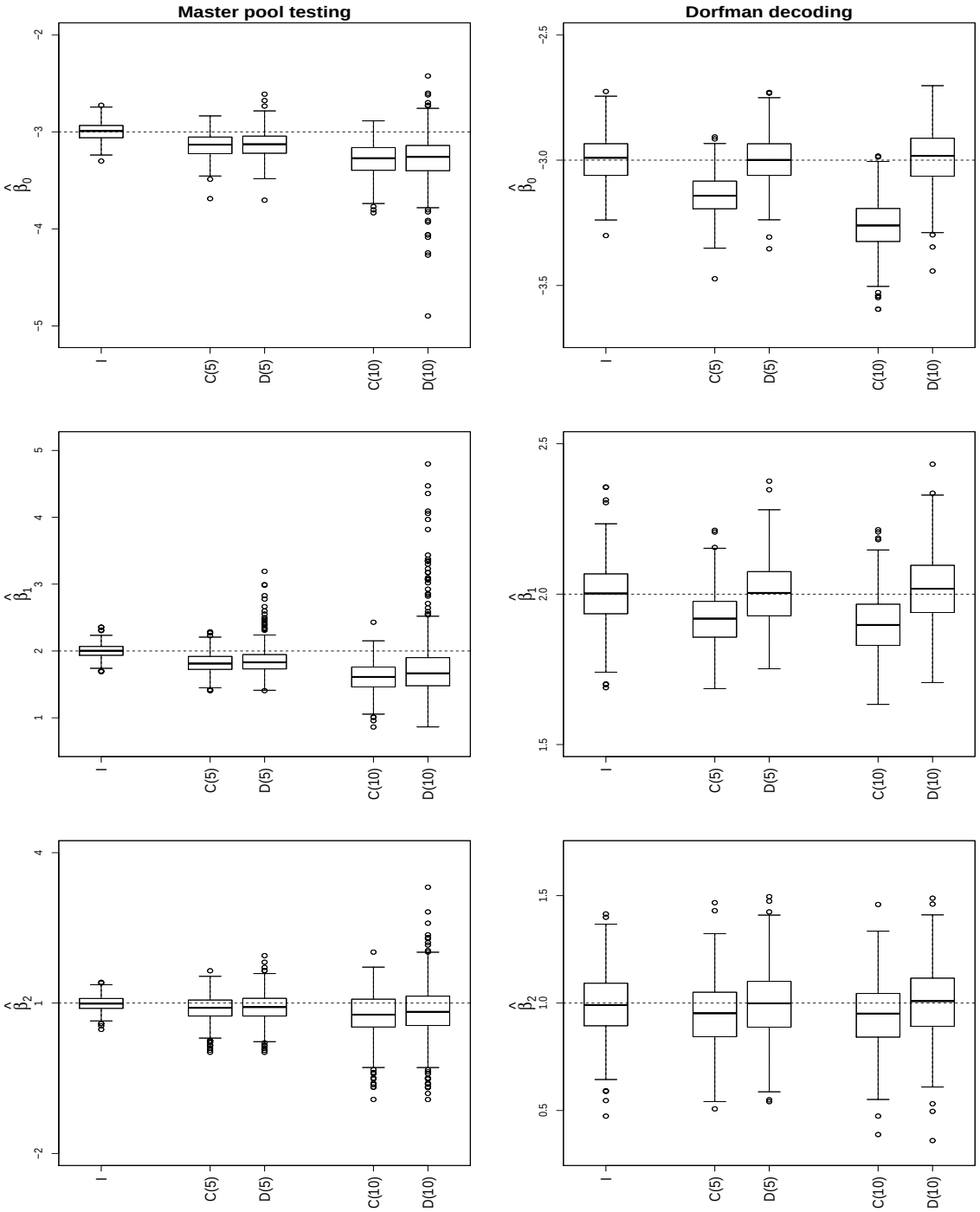

Figure B.1: Robustness study with misspecified submodel using **random pooling** and **moderate misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed method with the submodel in Equation (3.8) is fit to group testing data generated using the submodel 'HS'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

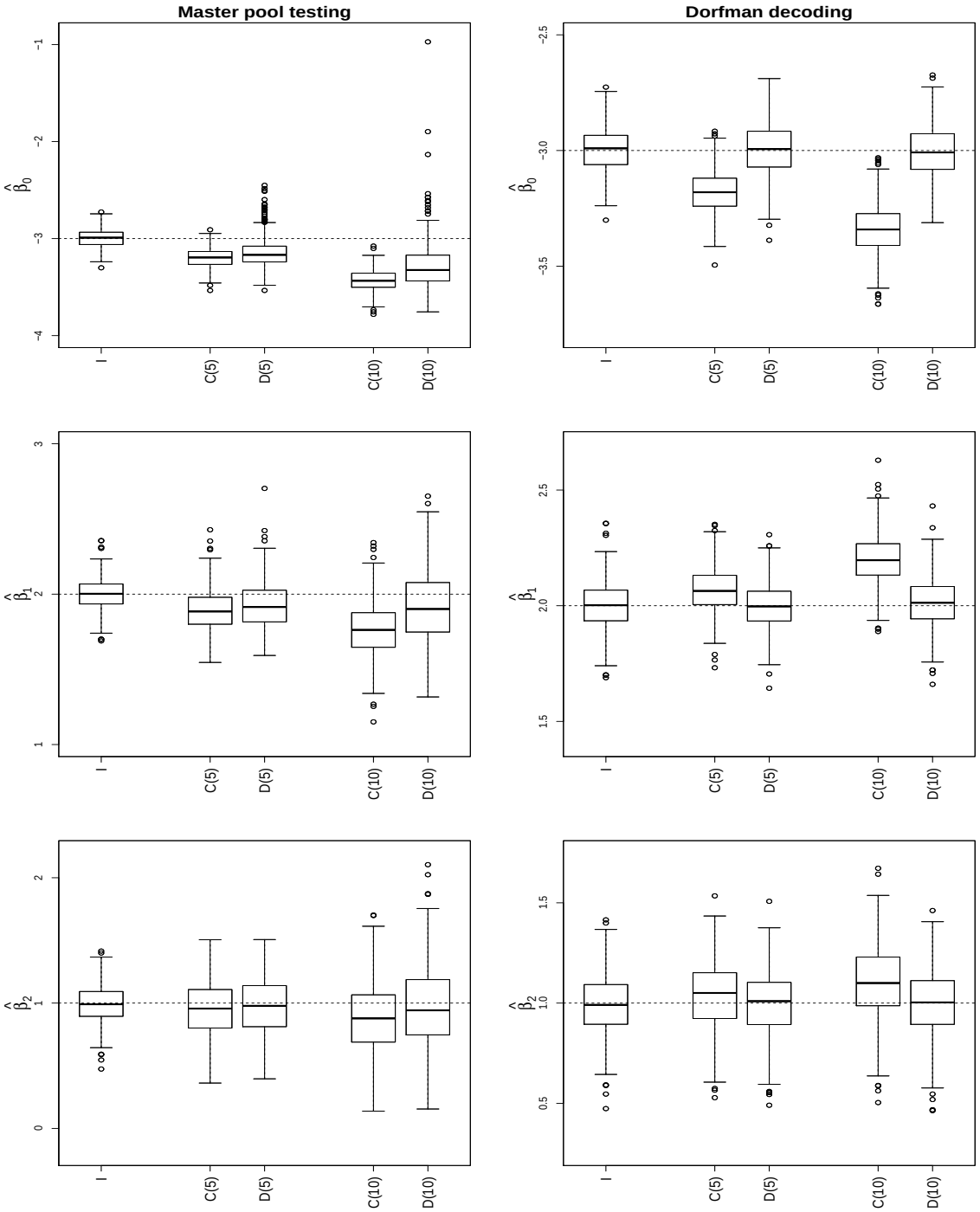

Figure B.2: Robustness study with misspecified submodel using **homogeneous pooling** and **moderate misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'HS'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

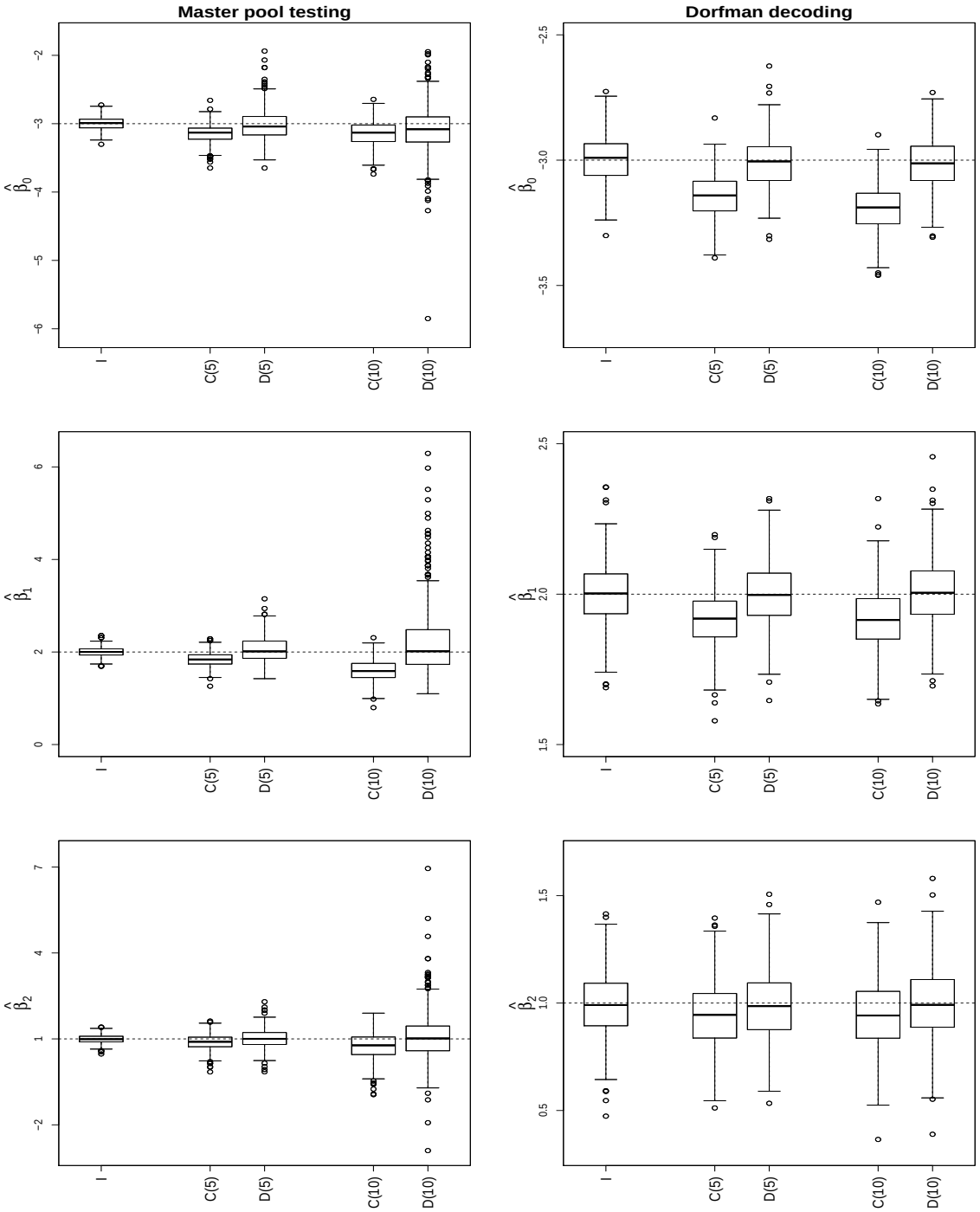

Figure B.3: Robustness study with misspecified submodel using **random pooling** and **moderate misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Probit'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

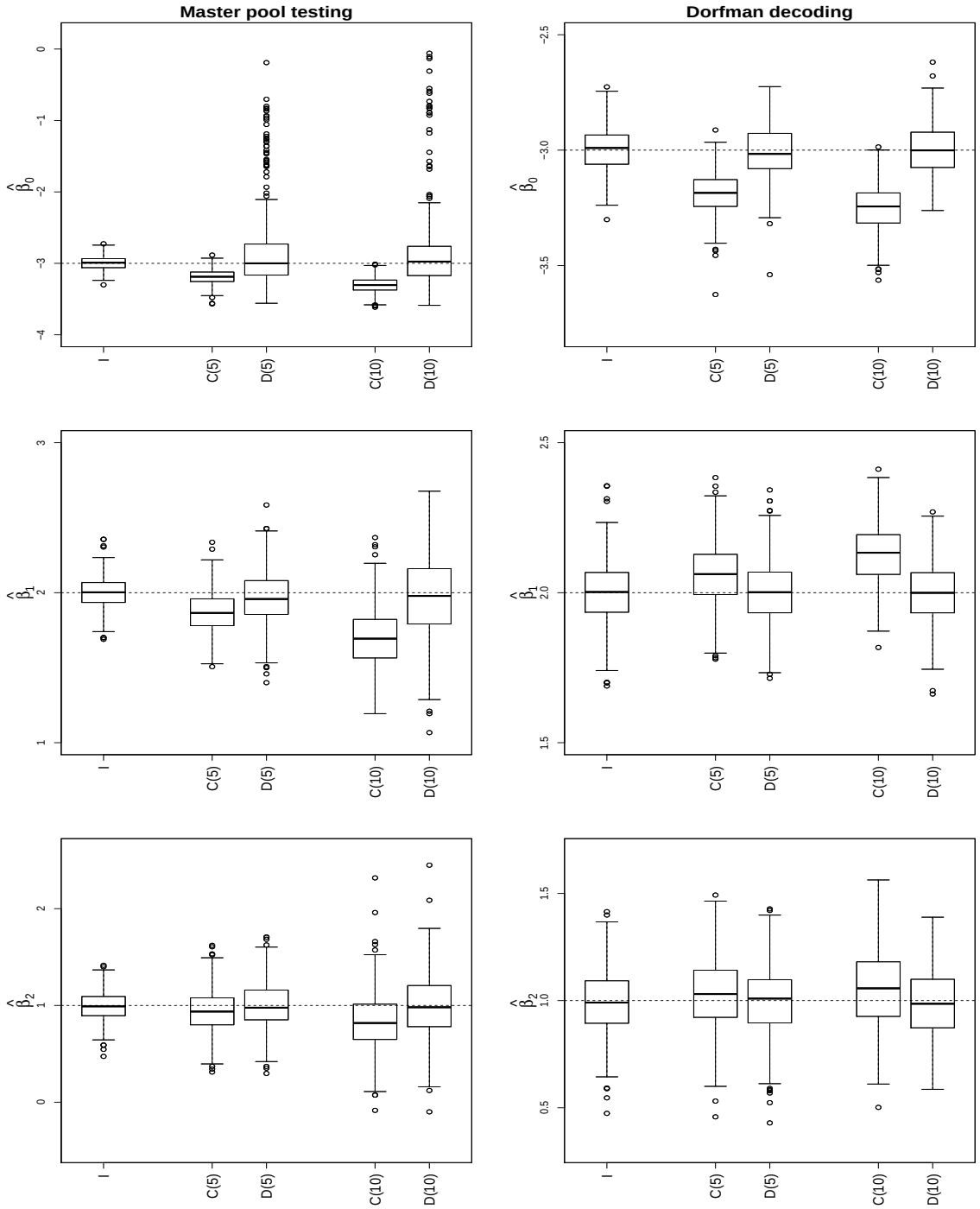

Figure B.4: Robustness study with misspecified submodel using **homogeneous pooling** and **moderate misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Probit'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

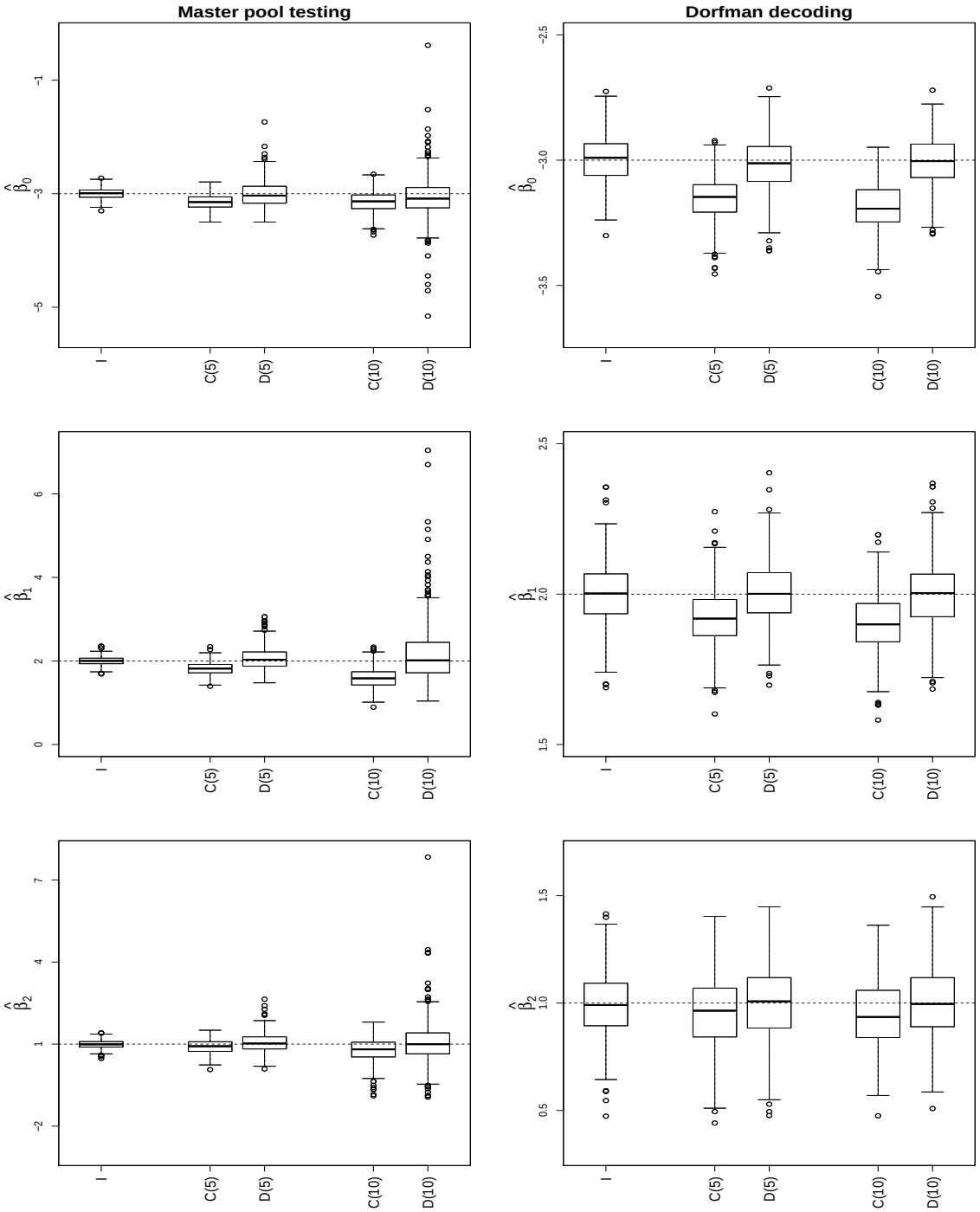

Figure B.5: Robustness study with misspecified submodel using **random pooling** and **moderate misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Cloglog'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

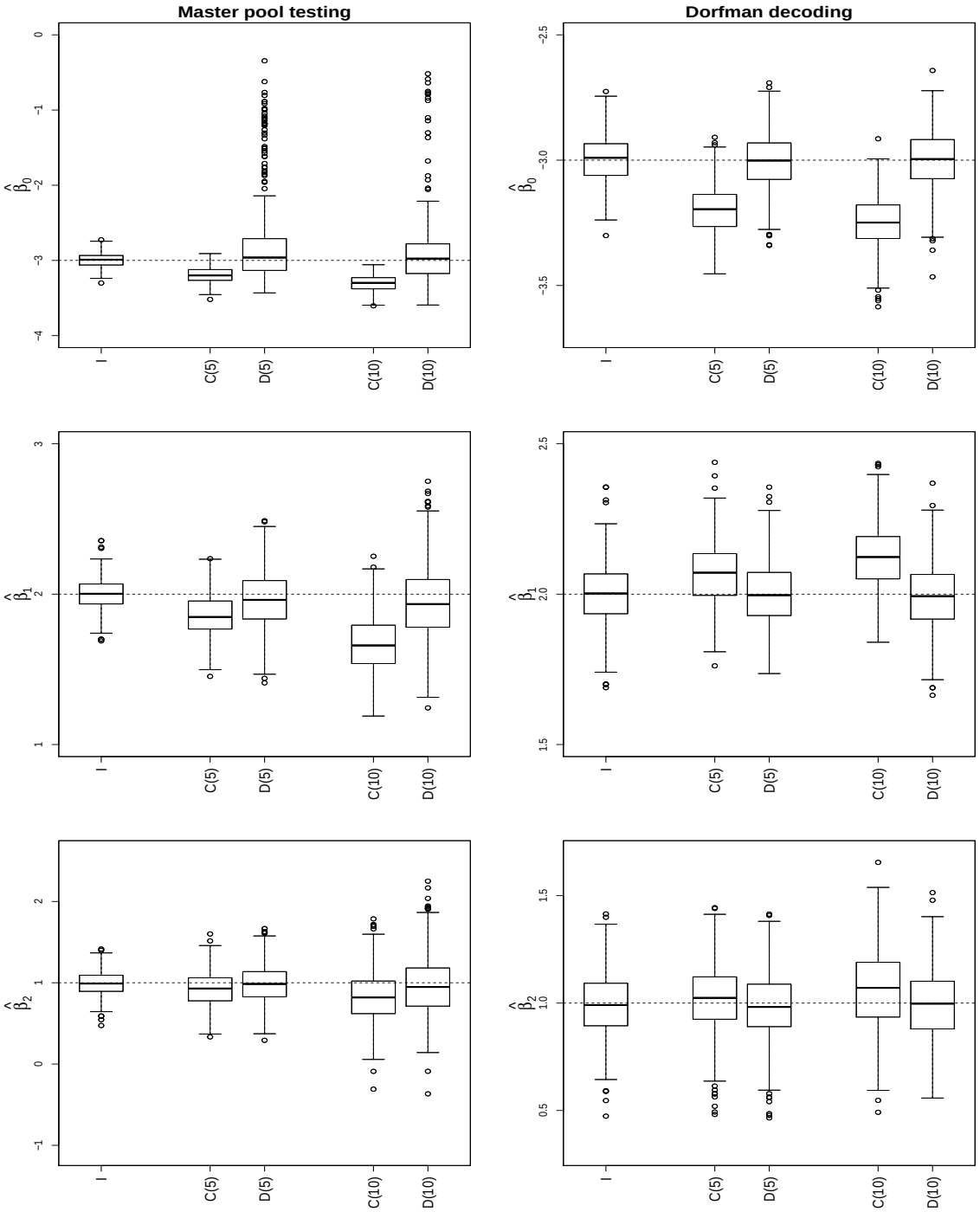

Figure B.6: Robustness study with misspecified submodel using **homogeneous pooling** and **moderate misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Cloglog'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

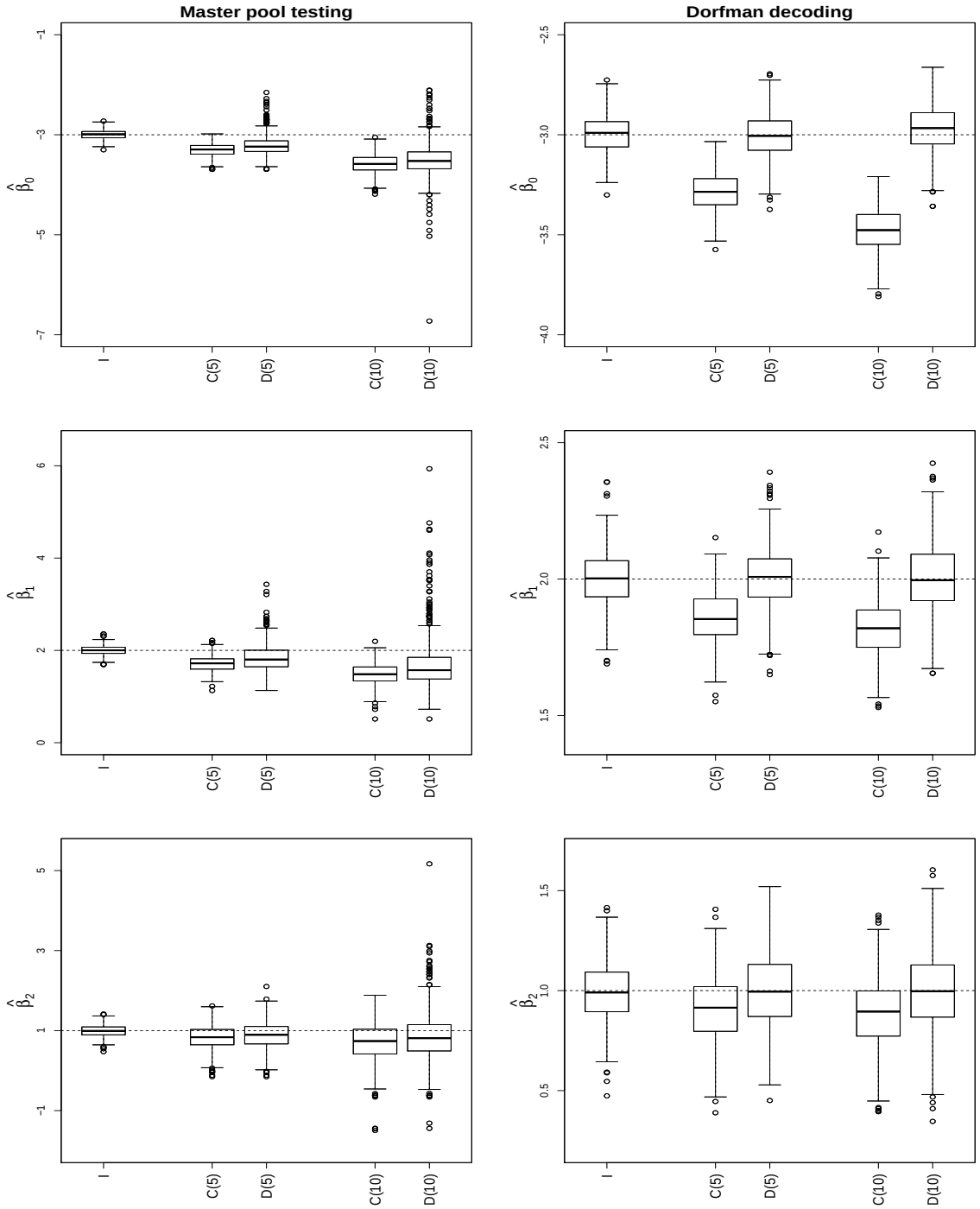

Figure B.7: Robustness study with misspecified submodel using **random pooling** and **severe misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'HS'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

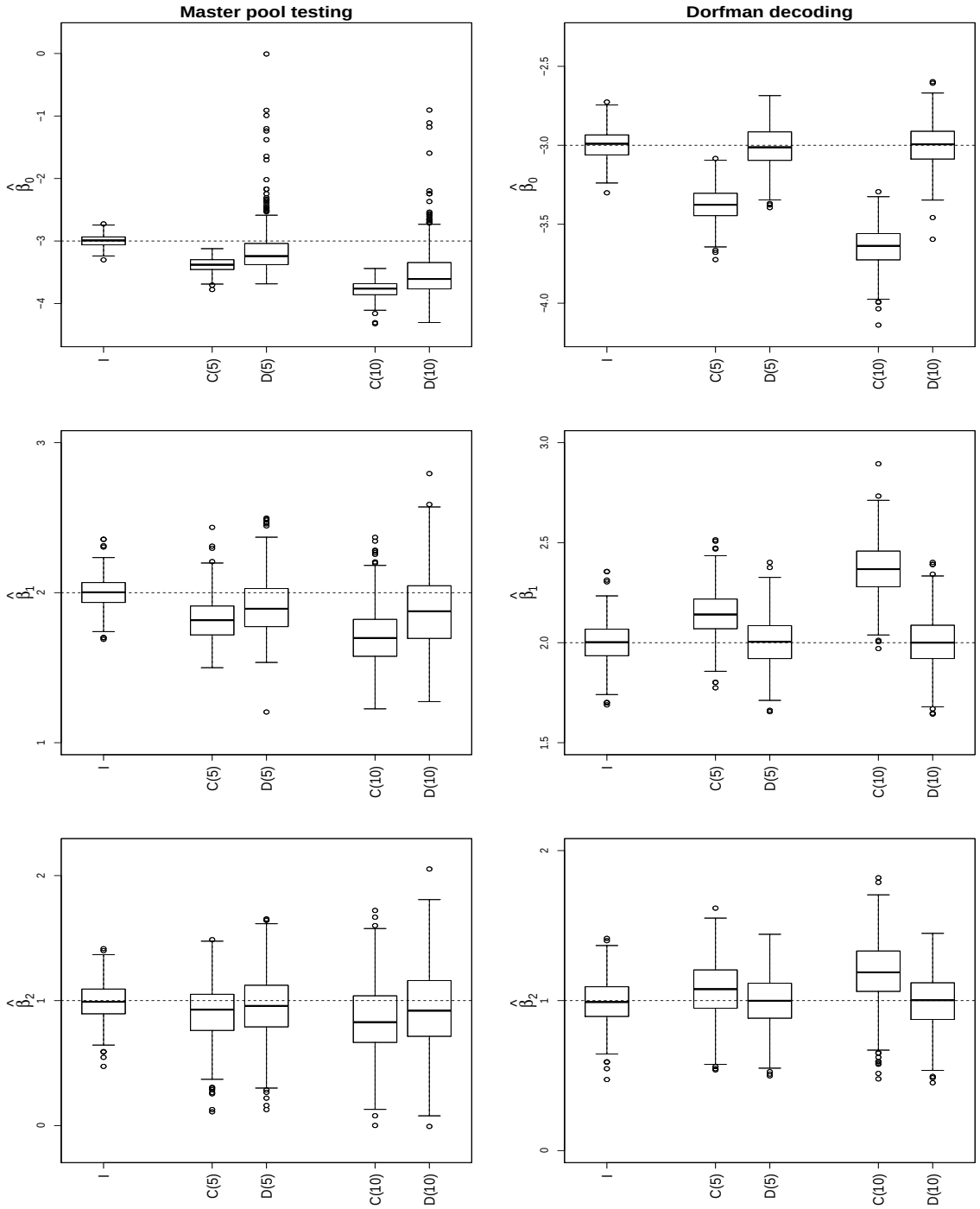

Figure B.8: Robustness study with misspecified submodel using **homogeneous pooling** and **severe misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'HS'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

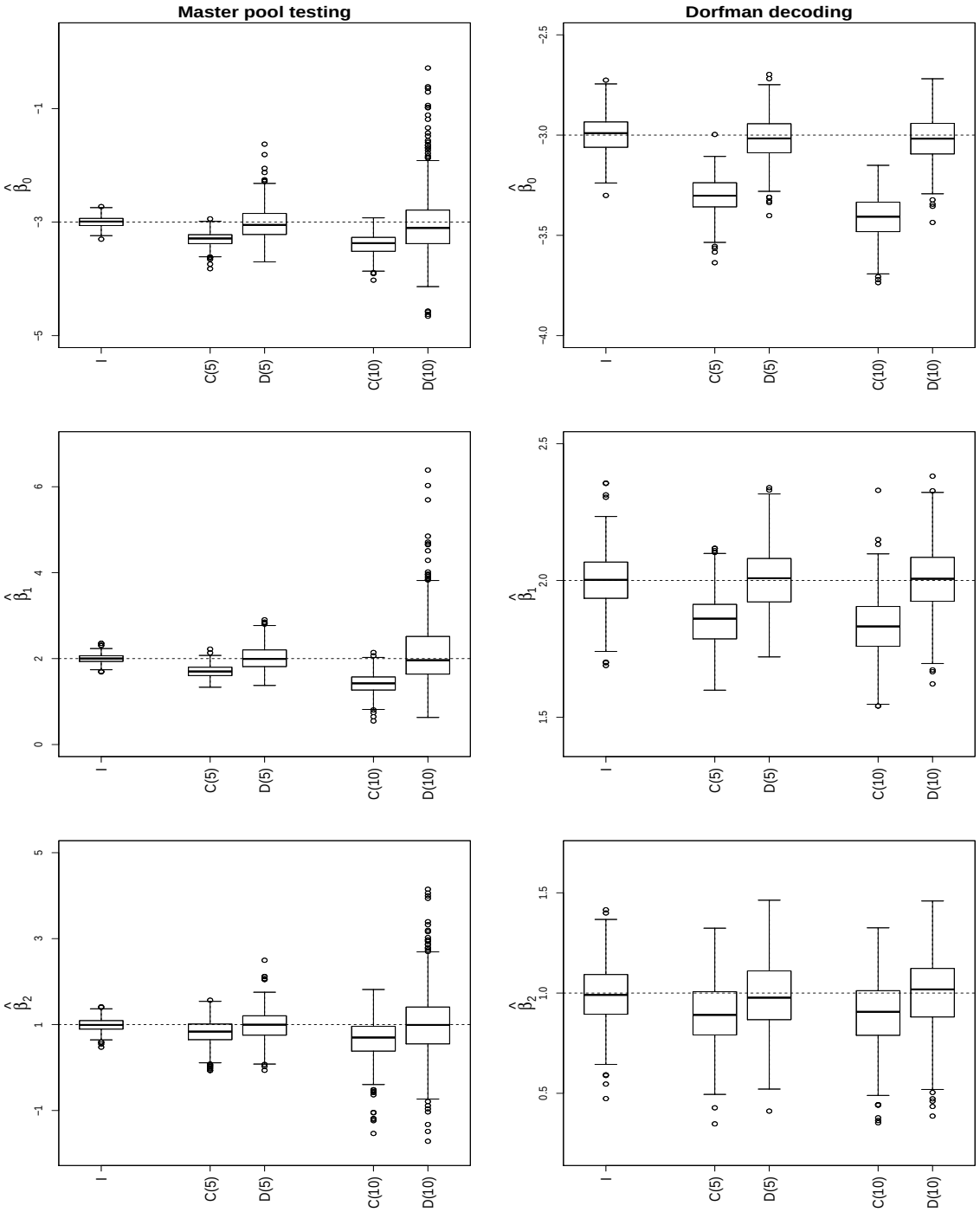

Figure B.9: Robustness study with misspecified submodel using **random pooling** and **severe misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Probit'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

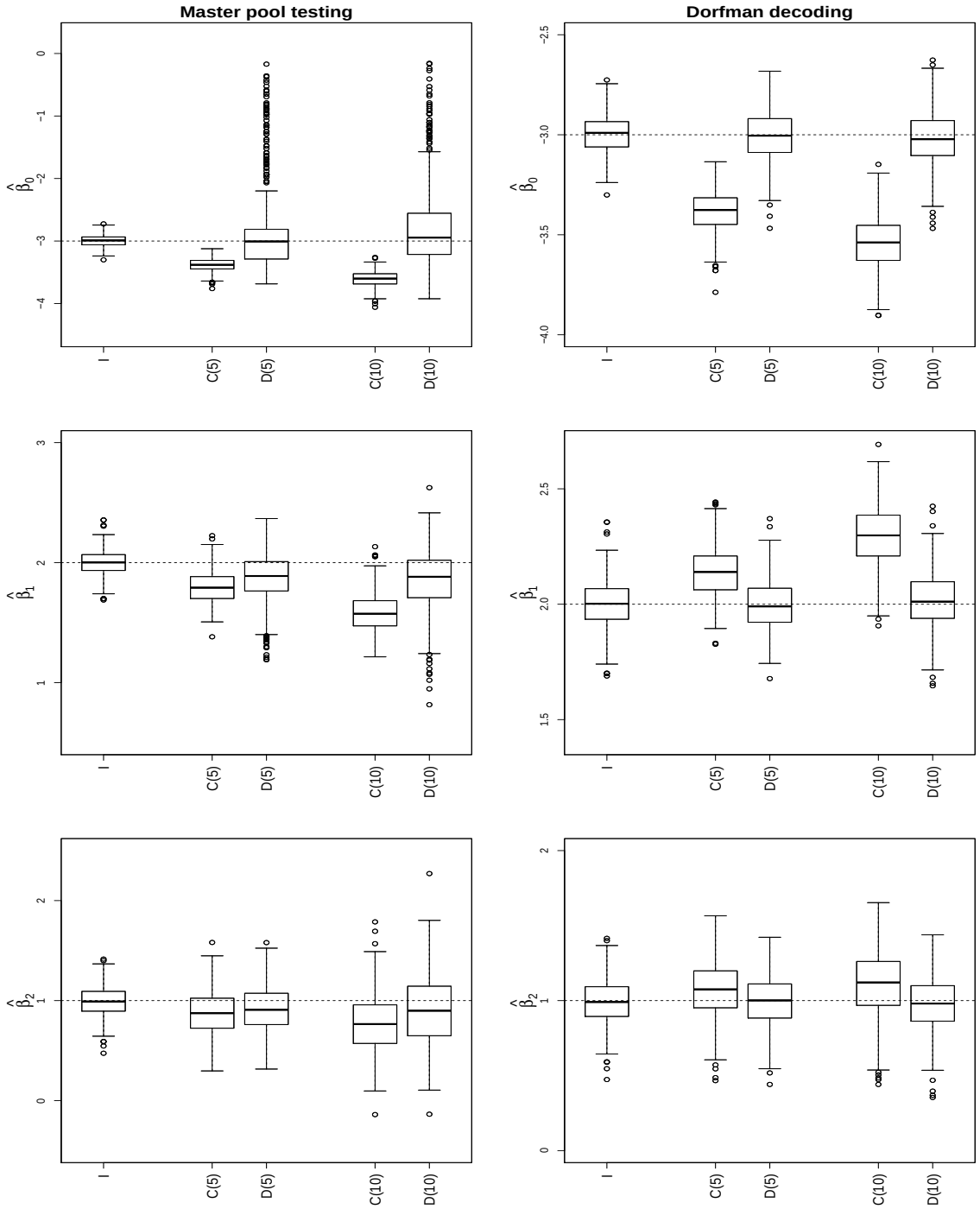

Figure B.10: Robustness study with misspecified submodel using **homogeneous pooling** and **severe misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Probit'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

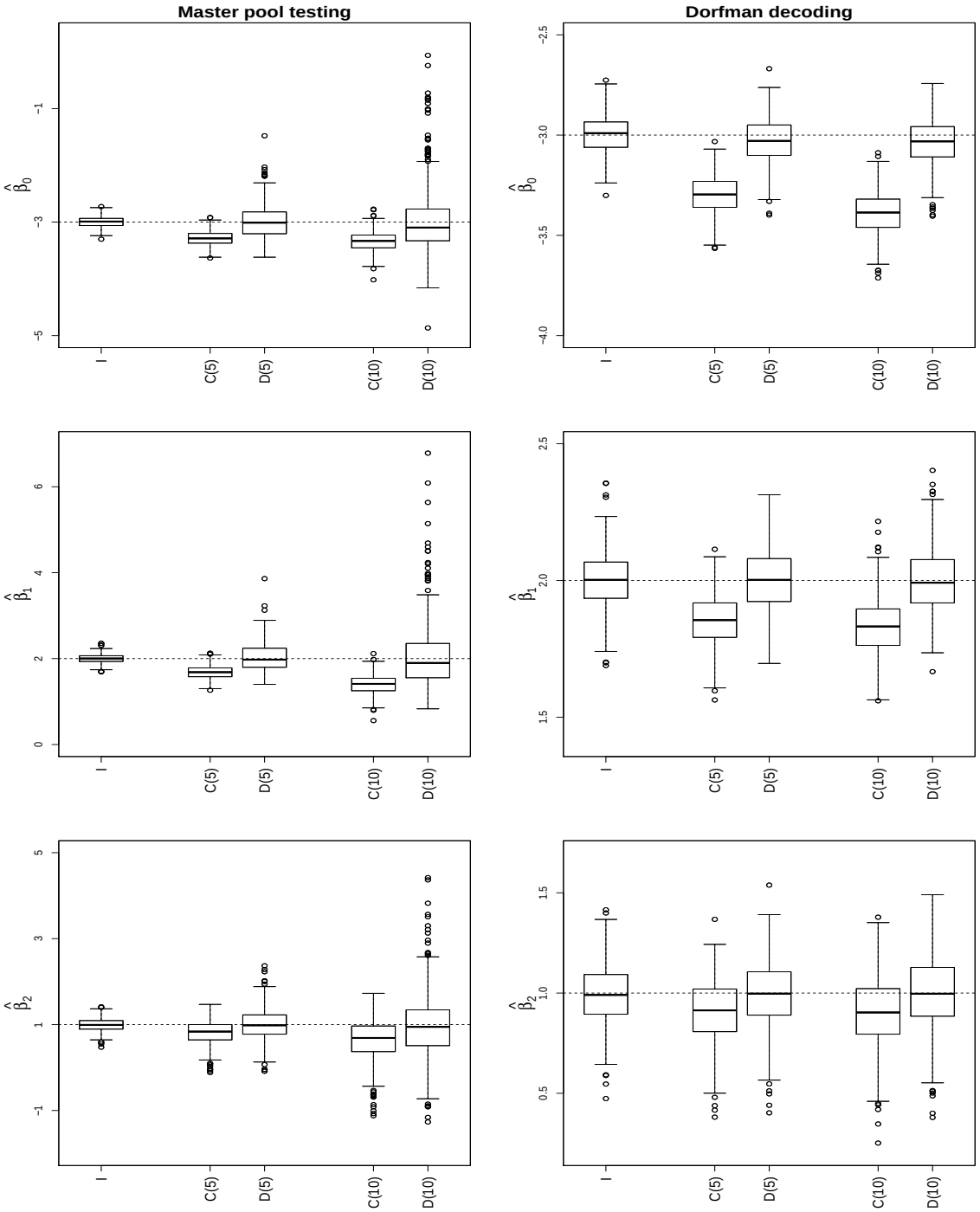

Figure B.11: Robustness study with misspecified submodel using **random pooling** and **severe misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Cloglog'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

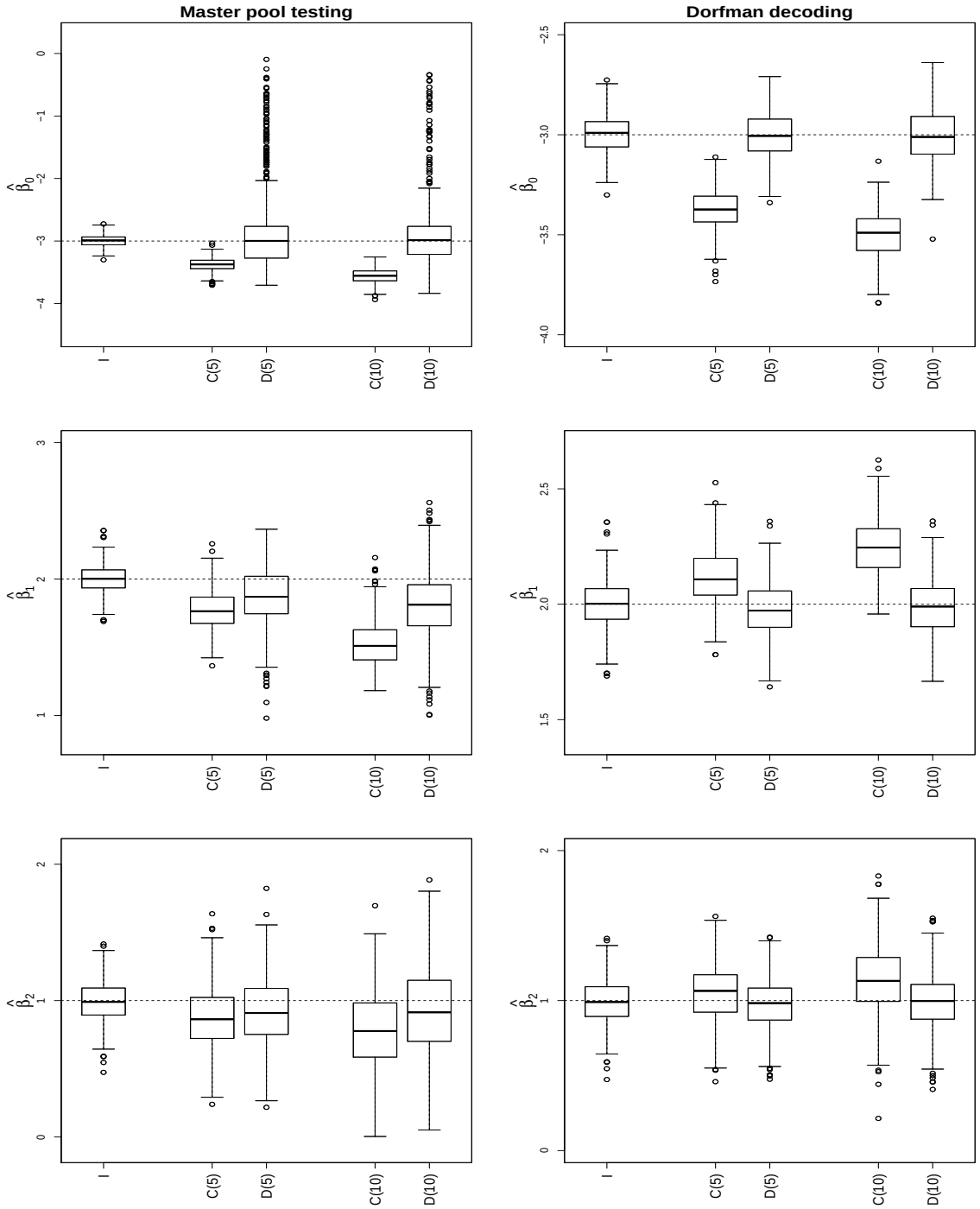

Figure B.12: Robustness study with misspecified submodel using **homogeneous pooling** and **severe misclassification**. Boxplots of the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ from 500 simulated data sets are presented. The proposed model with the submodel in (3.8) is fit to the group testing data generated using the submodel 'Cloglog'. On the horizontal axes, 'I' refers to individual testing, 'C' refers to the constant S_e/S_p method, and 'D' refers to our proposed dilution method. Constant pool sizes c are reported in parentheses. The true parameter values are represented by dashed horizontal lines.

B.6 The HBV data results from Section 3.5.

In this section, we present additional results of the HBV data application in Section 3.5.

- Table B.3: Estimation results for the polynomial model in Equation (3.10)
- Figure B.13: Estimated regression functions for the first order model in Equation (3.9)
  - homogeneous pooling is used
- Figure B.14: Estimated regression functions for the polynomial model in Equation (3.10)
  - homogeneous pooling is used

Table B.3: Irish HBV data analysis with Dorfman decoding. The polynomial logistic model in Equation (3.10) is assumed. MLE (estimated standard error) for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ averaged over $B = 500$ implementations. "Reject" is the proportion that the likelihood ratio test in Section 3.3 detects dilution using the level of significance α. Individual testing ($c = 1$) estimates are also reported for comparison.

	Constant S_e/S_p			Dilution			Reject	
c	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\alpha = 0.05$	0.10
1	−2.27 (0.12)	0.53 (0.16)	−0.15 (0.08)	—	—	—	—	—
				Random pooling				
3	−2.42 (0.14)	0.58 (0.18)	−0.21 (0.09)	−2.19 (0.21)	0.59 (0.18)	−0.22 (0.10)	0.35	0.43
4	−2.47 (0.14)	0.55 (0.18)	−0.20 (0.09)	−2.18 (0.21)	0.56 (0.18)	−0.20 (0.10)	0.50	0.66
5	−2.51 (0.14)	0.51 (0.18)	−0.18 (0.09)	−2.16 (0.20)	0.52 (0.18)	−0.18 (0.10)	0.72	0.80
6	−2.54 (0.14)	0.47 (0.18)	−0.16 (0.09)	−2.15 (0.19)	0.48 (0.18)	−0.17 (0.10)	0.81	0.89
8	−2.57 (0.14)	0.44 (0.18)	−0.15 (0.09)	−2.13 (0.18)	0.45 (0.18)	−0.15 (0.09)	0.94	0.98
10	−2.58 (0.14)	0.42 (0.18)	−0.14 (0.09)	−2.11 (0.18)	0.44 (0.18)	−0.14 (0.09)	0.99	0.99
				Homogeneous pooling				
3	−2.41 (0.14)	0.61 (0.18)	−0.23 (0.09)	−2.22 (0.21)	0.59 (0.18)	−0.23 (0.10)	0.28	0.38
4	−2.45 (0.14)	0.59 (0.18)	−0.22 (0.10)	−2.20 (0.20)	0.57 (0.18)	−0.22 (0.10)	0.43	0.58
5	−2.48 (0.14)	0.57 (0.18)	−0.21 (0.09)	−2.16 (0.19)	0.54 (0.19)	−0.21 (0.10)	0.71	0.82
6	−2.51 (0.14)	0.52 (0.18)	−0.19 (0.09)	−2.13 (0.19)	0.49 (0.19)	−0.18 (0.10)	0.87	0.95
8	−2.53 (0.14)	0.49 (0.18)	−0.18 (0.09)	−2.09 (0.18)	0.45 (0.19)	−0.17 (0.11)	0.98	0.99
10	−2.55 (0.14)	0.49 (0.18)	−0.14 (0.09)	−2.08 (0.18)	0.45 (0.19)	−0.15 (0.10)	1.00	1.00

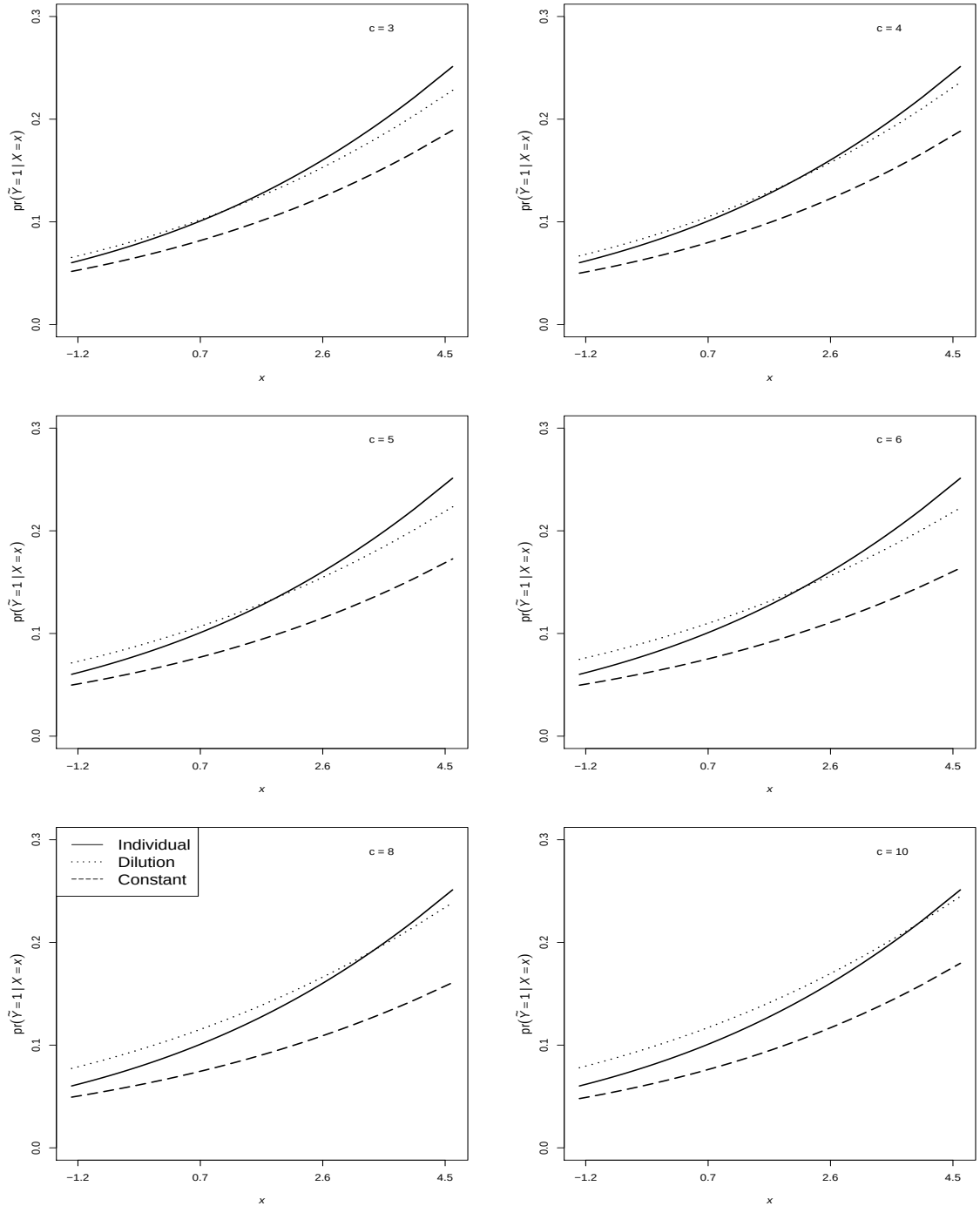

Figure B.13: Irish HBV data analysis with Dorfman decoding and homogeneous pooling. The first-order logistic model in Equation (3.9) is assumed. Estimated regression functions, averaged over $B = 500$ implementations, are presented. The estimated regression function for individual testing is also shown for comparison.

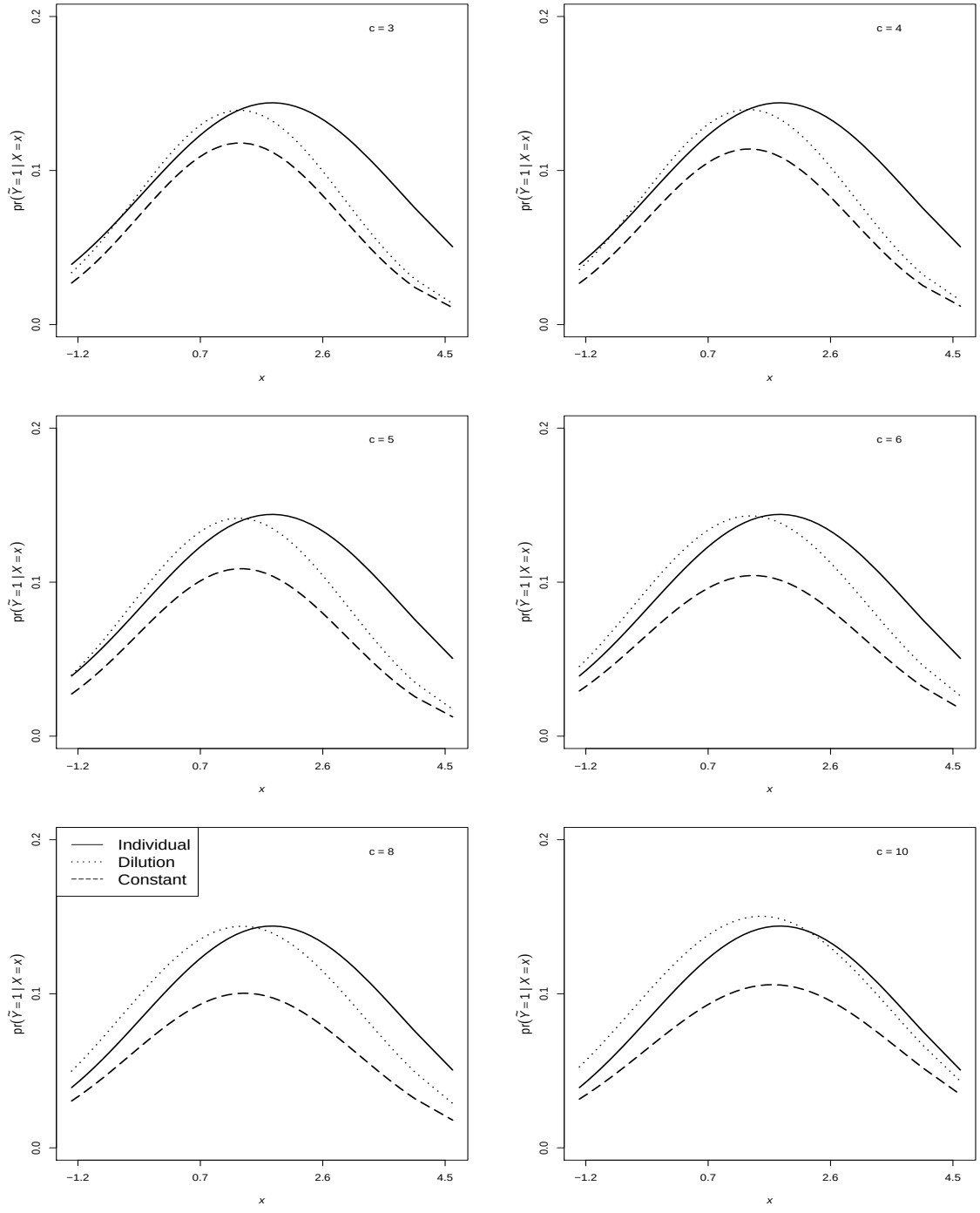

Figure B.14: Irish HBV data analysis with Dorfman decoding and homogeneous pooling. The polynomial logistic model in Equation (3.10) is assumed. Estimated regression functions, averaged over $B = 500$ implementations, are presented. The estimated regression function for individual testing is also shown for comparison.

Appendix C

Permission to reprint

This section shows the evidence that the author of this dissertation has permission to reprint the material of the article, "Estimating the prevalence of multiple diseases from two-stage hierarchical pooling," presented in Chapter 2 and in Appendix A. Note, the legal name of this author is 'Md Shamim Sarker'. However, the author uses the name 'Md S. Warasi' for publications and conferences as it appears on the article.